Controlling Synthetic Biological Systems at the Plasmid and Population Levels

Alexander Joseph Harper Fedorec

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Centre for Computation, Mathematics and Physics in the Life Sciences and

Experimental Biology

University College London

December 4, 2018

I, Alexander Joseph Harper Fedorec, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

The tools and methods for manipulating biological systems have been advancing at a great pace for the last few decades. These advances make possible the development of ever more sophisticated synthetic biological circuits for varied applications; from the bio-remediation of polluted environments to the dynamic delivery of targeted therapeutics in the body. However, the burden placed on host cells carrying these synthetic circuits leads a reduction in fitness compared to the unengineered host or wildtype environmental strain. This leads to challenges in maintaining a functioning circuit within a host population and ensuring host population existence in a competitive environment. In this thesis, I explore - and engineer - stability of synthetic biological systems at the circuit and population levels.

In the first part of the thesis, I consider the population levels of multi-species bacterial communities within a chemostat environment. The difficulty with constructing such communities is that, in the absence of other interactions, differences in growth rates lead to competitive exclusion of the slower growing strain(s). I design and construct a two-species community that circumvents competitive exclusion. In this system, the slower growing strain is able to dynamically relieve that pressure by producing a bacteriocin when its population drops too low; the first system to use bacteriocins in this way. Using a mathematical model and computational simulation I show the different achievable community dynamics and demonstrate that they are easily experimentally tunable using an inducer molecule or dilution rate. The novelty of this design is that it can produce stable multi-species communities while only requiring the engineering of one of the constituents. The full system is constructed and characterised using flow cytometry and Bayesian param-

eter inference. Finally, the system is tested in competition in a bioreactor and shown to extend the presence of the two species.

In the second part of the thesis, two approaches to enhancing the stability of plasmids in a bacterial host are demonstrated: toxin-antitoxin (TA) systems and bacteriocins. This is the first characterisation of axe/txe and microcin-V in the commensal bacteria *E. coli* Nissle 1917; a strain commonly used in the development of engineered probiotics. I show that both approaches improve the stability of burdensome plasmids *in vitro* and in an *in vivo* mouse tumour model. The bacteriocin is shown to have the added ability of being able to reconstitute a plasmid-bearing population if it is invaded by competitors. Novel deterministic mathematical models and Bayesian statistical methods are used to understand and explain the differences in the performance of these systems.

Systems such as these will enable the deployment of synthetic biological systems in to ever more varied environments, opening doors to a great number of future applications.

Impact Statement

The research presented here may prove useful within academia and industry. Two new systems are explored for tackling plasmid stability; a fundamental problem in the deployment of synthetic biological systems in non-laboratory environments. These systems are of particular interest for use in industrial bioreactors due to the current cost of the removal of antibiotics from any products that are being produced. Further, these systems were demonstrated in the commonly used probiotic bacterial strain *Escherichia coli* Nissle 1917 and their function demonstrated in an *in vivo* setting. This shows a clear application with regards to the development of future live biotherapeutics. I have also demonstrated a new approach to multi-species population control. This is another challenge that is being faced by industry as microbial communities rather than single strain populations of engineered microbes become more relevant.

Acknowledgements

There are far more people to thank than I will write about in this section. This, and I suspect any, thesis is only possible thanks to the advice, support and friendship of a great number of people over many years. Saying that, I must start by acknowledging my supervisor, Dr Christopher Barnes. He gave me freedom, encouragement and support to explore and pursue more ideas than are just presented here. This, of course, was always balanced with a healthy dose of cynicism. I am so thankful to have been a part of his research group as it has grown and explores ever more exciting ground. Secondly I want to thank my fellow PhD student, and now Doctor, Tanel Ozdemir. He welcomed me in to the lab, patiently taught me, pushed me every step of the way and has become an incredible friend. Without him this whole process would certainly have been less enjoyable and less successful. I would also like to thank Professor Geraint Thomas for his guidance and support as my secondary supervisor, Dr Miriam Leon for being the most welcoming face in all of London, and all of the people that have been a part of the Computational Systems and Synthetic Biology group. Finally, from UCL, I would like to thank all of the students and staff in CoMPLEX who have created an eclectic, nurturing community of young researchers. Particularly, I would like to acknowledge the tight-knit group of friends that come out of the 2013 intake.

Before I came to UCL, there were obviously several people who played important roles in directing me towards scientific research. I suspect the most significant among them is my undergraduate dissertation supervisor, Dr Joanna Bryson. It had not crossed my mind that I was capable of undertaking a PhD until I worked with her. She is my current model for an exceptional interdisciplinary researcher. My

personal tutor, Professor James Davenport, not only helped me to get through my undergraduate degree but was integral in getting me onto the program in the first place. I would like to thank him for sharing his experiences, about far more than academia, with me. I would also like to mention Dr Alan Hayes, Dr Leon Watts and Dr John Power for their support throughout my four years in the Department of Computer Science.

In the four (and a bit) years that it has taken me to complete this PhD, my partner Sarah has proposed to me and together we have shared the most magical wedding day. She has been amazingly supportive throughout my studies and has rarely complained about the late nights or lost weekends. Although I would often be too tired to talk about my research with her at the end of the day, she would always try and catch any snippets of information that she could. Finally I would like to thank my family. My dad, Alex, has been my main source of support when I have struggled and has helped me immensely to reach this point. It seems that no matter where I tread, I am following in his footsteps. I am insanely jealous of my sister, Katie's, capacity for creativity. It is certainly unfortunate for any reader of this thesis that I don't share more of her abilities. Finally, I would like to dedicate this thesis to my mum, Mary. Although she didn't live to see me reach this point, she lives on in my sister and I. She wouldn't be surprised that I've made it here.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Modern DNA manipulation techniques have allowed researchers to design and implement gene regulatory networks for therapeutic, industrial and research purposes in both prokaryotic and eukaryotic organisms. From the production of valuable compounds, such as artemisinin, in bioreactors to the targeting of genetic diseases using phage therapy, engineered biological systems are already being used outside of the laboratory and their development and deployment will continue apace. In this Chapter I will give a brief overview of the field of synthetic biology and two current challenges in the field that this thesis will address.

1.1 Overview

The discovery of DNA and the elucidation of the flow of information in molecular biology, as shown in Figure 1.1, has allowed researchers to think of biological organisms as systems that are able to be engineered. The development of methods for the cutting and splicing of DNA enabled rudimentary changes to organisms through the removal or addition of genes. More recently, genetic engineering has advanced from knocking out individual genes in an organism to a point where it is now tractable to create new systems of multiple interacting genes or synthetic circuits [2]. Engineering principles, such as abstraction and modularity [3], and methodologies, such as the design-build-test cycle, have been applied to the design and implementation of these complex systems to produce the field of synthetic biology. In fact synthetic biology now extends beyond the piecing together of ex-

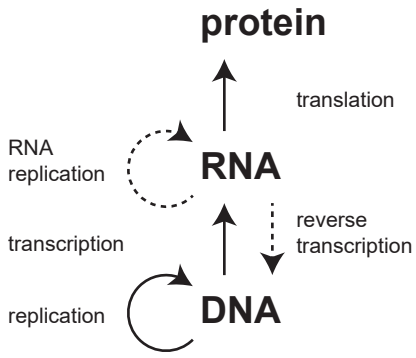

Figure 1.1: The flow of information in a biological system, often called the central dogma. The dashed lines represent uncommon flows.

isting biological parts to the design of entirely new genes, proteins and even DNA analogues.

1.1.1 Applications of Synthetic Biological Systems

Synthetic circuits in host cells, such as bacteria, can take inputs from the environment and the host and produce outputs that can report on and alter the current state of the environment or host. Cell-free systems are now being developed that remove the host cell from the system and create direct interactions between the environment and the circuit [4]. A library of modules and components have been developed over the last 20 years including many analogues of electronic components such as Boolean logic gates [5], memory [6], and clocks [7]. As our understanding of interactions between circuit components and host cells improves, and technologies for the building of circuits expand, we are able to integrate these components in to ever more sophisticated devices.

In an industrial setting 'biological factories' are being produced that enable the production of products ranging from simple hydrocarbons such as bioethanol [8] through to complex pharmaceuticals [9]. Organisms are being developed for the bioremediation of environmental pollutants such as arsenic [10]. Synthetic biology also provides tools for further biological research, with optogenetics enabling the reporting or triggering of neuron firing [11].

One of the most interesting and challenging sectors in which synthetic biology

is being used is in the development of live biotherapeutics. Systemically delivered drugs are untargeted, leading to only a small fraction of the therapeutic reaching the intended location with significant off-target delivery giving rise to unwanted side-effects. Temporal control of drug activity is reliant on dosing regimens with predicted systemic circulation and metabolism. An alternative approach is to leverage the tools of synthetic biology to develop microbes that can perform targeted therapeutic delivery with much greater control of location and timing, ultimately allowing for higher therapeutic doses with lower systemic effects [12]. Figure 1.2 shows how engineered bacteria can be used in the human gastrointestinal tract to target potential disease states. Recent examples of engineered bacteria being used for therapeutics include the detection of liver cancer [13] and the prevention of *Pseudomonas aeruginosa* infection [14]. In order to make bacterial therapeutics a reality for humans there are a number of challenges still to face. Creating circuits that function predictably in any of the potential environments that they are exposed to is a primary necessity. Further, concerns of *de novo* genetic constructs escaping from their intended host organisms must be tackled.

1.1.2 Synthetic Circuits in Host Organisms

Bacteria provide a tractable host for such synthetic biological circuits for several reasons but none less than due to our level of understanding of the inner workings of a set of model bacterial organisms and ongoing developments in expanding this set. A host bacterium can carry the genes comprising a synthetic circuit in two ways: on its chromosome(s) or on extra-chromosomal material such as pieces of circular, double stranded DNA called plasmids. There are arguments for and against each method and both have been used successfully in a variety of organisms.

By encoding a synthetic circuit on the chromosomal DNA of the desired organism one ensures the stable inheritance of that circuit. Further, one can be confident of the copy-number of the circuit within the host, leading to potentially greater predictability of circuit behaviour. Inserting genetic circuits into the chromosome of bacterial species is not necessarily simple, though new methods using lambda-Red [16] and CRISPR-Cas9 [17] are making the process easier. However, bacteria are

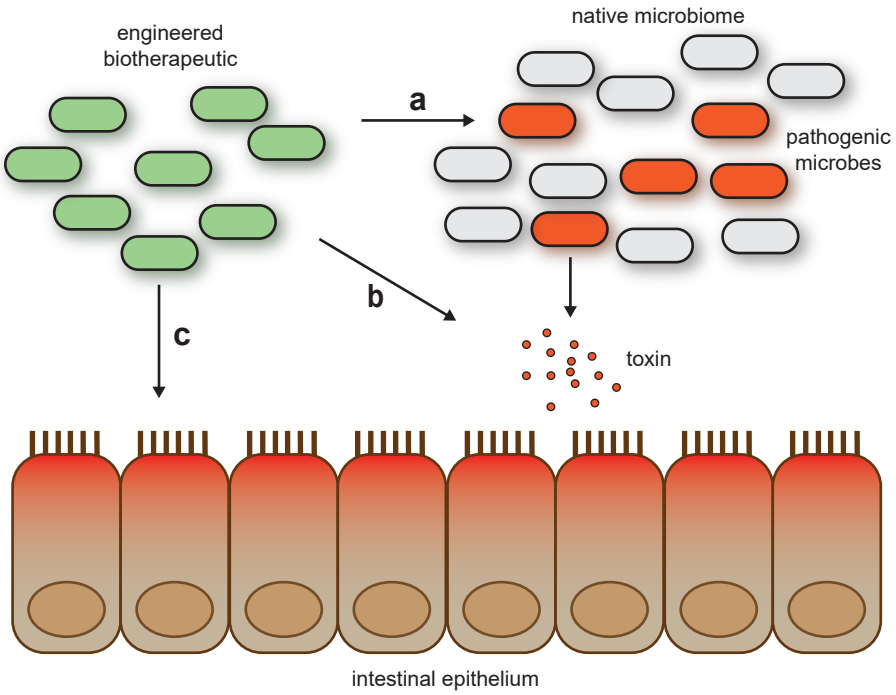

Figure 1.2: The gastrointestinal tract is an example of an environment in which live biotherapeutics can provide targeted benefit. Engineered strains have been designed to target (a) pathogenic microbes, (b) molecules that may lead to pathological condition, and (c) specific disease states. Figure taken from [15].

good at streamlining their chromosomes [18]. Although there are a large number of nonessential genes in the *E. coli* genome [19], it has been shown that the spatial positioning of a gene within a cell affects its rate of transcription [20, 21]. This may lead to unpredictable behaviour unless the circuit is placed in a well characterised location. Finally, it is common for the development of circuits to be carried out using plasmids due to the ease of their manipulation but recent work has shown that moving from multi-copy plasmids to single copy chromosomal insertions can lead to unexpected change in circuit function [22].

The use of plasmids as vectors for synthetic circuits is common. Plasmids are a fundamental biological tool and have been widely used in molecular and cellular biology research, leading to a number of well developed methods for their manipulation [23, 24]. This ease of manipulation enables a level of modularity which is one of the key engineering goals of synthetic biology [25, 3]. Plasmid copy number can be an important parameter in the functioning of circuits, and it is non-trivial

to convert a dynamical gene circuit from a multi-copy plasmid implementation to a single copy implementation and maintain identical function [22]. As such, plasmids remain the main vehicle for genetic circuits in industrial batch processes.

However, there are a number of difficulties with their use. Basic problems with the use of plasmids exist such as plasmid incompatibility [26, 27] and unwanted horizontal gene transfer [28], though these can often be overcome in the design. More fundamental is the problem of stable maintenance of plasmids across multiple generations. As synthetic plasmids often solely carry non-essential genes, their loss when a cell divides leads to an advantage for the daughter cell that does not possess the plasmid [29]. I will outline this problem in detail in this Chapter and demonstrate two different types of system to solve this problem in Chapter 5.

1.1.3 Synthetic Populations in Competitive Environments

As synthetic biological devices begin to leave the laboratory and are used in more complex environments the challenge becomes not just maintaining a functioning circuit within a host cell but maintaining a functioning population of cells within an environment. To do this we must consider how the systems that we design interact with competitors in the environment. To be able to design systems that will function in such environments we must be able to predict, and ideally control, the residency of the cells that make up the systems. A system that produces a targeted therapeutic is not useful if its occupation of the target environment is so transient that the therapeutic is not produced. Equally, we must take care to minimally disturb the present inhabitants of such environments, unless that is the goal of the system, as healthy ecosystems often require the balance of many species [30].

In addition to the challenges of maintaining a single synthetic population in a complex environment, systems are now being designed which take advantage of complex heterogeneous synthetic populations. The burden that increasingly more complex circuits place on host cells is leading some researchers to develop systems in which there is a division of labour [31]. Instead of having a single cell produce every protein needed in a complex pathway or perform every computation in a logic network, processes can be distributed between a heterogeneous population of cells,

each capable of undertaking different tasks, passing metabolites between them as necessary and communicating through small molecules. By instituting a division of labour, one can reduce the burden on individual cells, reducing the fitness cost and the evolutionary pressure to gain mutations which reduce the burden, and the function, of the circuit [32].

In Chapter 3 I will describe a system that I have developed for the predictable control of an engineered population in an environment with competition. In Chapter 4 I will describe the construction and characterisation of such a system. This system enables the residency of two populations with different growth rates, combating competitive exclusion, while only necessitating the engineering of a single strain. This affords its use in competition with wild-type strains or previously engineered systems.

For the remainder of this Chapter I will detail how the dynamic behaviour of plasmids, though potentially beneficial to bacteria and plasmids alike in nature, cause problems for their use as vectors in synthetic biology. In a laboratory setting segregational instability is often disregarded as antibiotic selection, through the expression of antibiotic resistance genes from the plasmids, is an easy way to ensure plasmid-free cells are killed. However, with the expansion of the environments in which synthetic biological systems are used, other solutions are necessary. The cost of removing and deactivating antibiotics from industrial fermentors means that they are not used [33]. For clinical applications, it has been shown that antibiotics increase horizontal gene transfer of resistance genes [34, 35] and it is likely that disruption of the native microbiota would be undesired [36]. As such, there have been efforts to use existing stability mechanisms such as toxin-antitoxin (TA) systems [37, 13], active-partitioning mechanisms [13, 38] and auxotrophy [39, 40] to improve plasmid maintenance.

Further, I will discuss how, as the complexity of the system being designed and built increases, synthetic biologists are beginning to spread the burden of systems across communities of cells rather than expecting a single strain to do everything. Constructing these synthetic communities produce their own challenges but also

open the door to new applications.

Finally, I will discuss some of the methods used to model biological systems in the field and in this thesis. I will also detail some Bayesian approaches to using these models to gain insight into experimental systems.

1.2 Plasmids Dynamics

Plasmids are extra-chromosomal pieces of double stranded, usually circular, DNA ranging in size from 1 kilo-base pair up to more than 1 mega-base pair [41]. They are most common in bacteria but have been found in other domains of life and are often used as vectors for the transfection of eukaryotic cells. The chromosomal DNA of bacteria has been compared to a poorly stocked home bookshelf, whereas plasmids give bacteria access to a vast library of genes [42, p. 24]. "It may take longer to obtain a book from the library than from one's own shelves but the library offers an enormous supply of titles at very low cost to the individual." A bacterium is able to gain possession of extra-chromosomal genes either through inheritance or through horizontal gene transfer, as shown in Figure 1.3. The ability to pick up these extra-chromosomal genes from the environment, pass the useful ones on to "offspring" and dispense with the burdensome ones, enables genetic experimentation to be separated from essential housekeeping chromosomal genes.

1.2.1 Horizontal Gene Transfer

Horizontal gene transfer enables bacteria to pick up genetic material throughout their lifetime rather than just through inheritance. The mechanisms through which this can occur are shown in Figure 1.3. Conjugation can carry both plasmids and conjugative transposons from one cell to another, with the latter capable of enabling the transfer of the former [44]. The GC content of mobile genetic elements tends to be lower than that of their hosts [45] which allows for the detection of recently inserted elements in the genomes of bacteria [46]. Transfer of plasmids by conjugation requires a plasmid-borne origin-of-transfer and a mobilisation mechanism which may be carried on the plasmid, on another extra-chromosomal genetic element or in the bacterial genome [47].

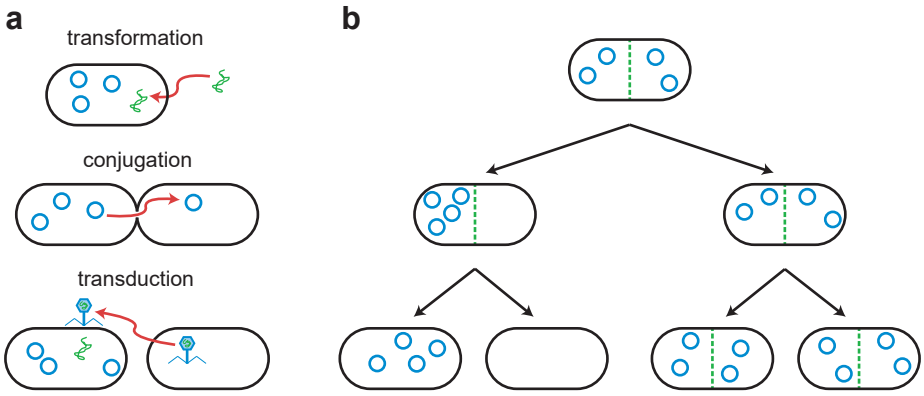

Figure 1.3: (a) Horizontal gene transfer can occur either through the acquisition of genetic material from the environment (transformation), the passing of genetic material from one cell to another (conjugation) or through infection by a virus (transduction). (b) Vertical gene transfer is the inheritance of genetic material by daughter cells when a cell divides. If the genetic material is not in both parts of the dividing cell, one of the daughters will be no longer carry those genes. Panel (b) is adapted from [43].

The spread of antibiotic resistance genes occurs through horizontal gene transfer and as such is a subject of much research. Carrying certain types of antibiotic resistance genes comes with a cost and, therefore, in environments not requiring resistance there is selection pressure to lose those genes [48, 49]. The expression of antibiotic resistance genes can be under transcriptional and translational regulation [50]. Plasmid-borne antibiotic resistance genes are often found on low copy plasmids [51] with the ability to increase copy number [52] and conjugation rate [53] in response to stress. It has been shown that in an environment in which the resistance genes are necessary, mutations can arise which counter the fitness costs [54].

Horizontal gene transfer is often considered a problem in synthetic biological applications as it may enable engineered genetic material to escape from its intended environment with potentially unknown consequences. This has led to systems being developed to prevent the replication or function of DNA outside of its intended setting [28, 55].

1.2.2 Vertical Gene Transfer

Vertical gene transfer is conceptually rather simpler than horizontal gene transfer. It is the inheritance of intra-cellular genetic material by the daughter cells of a dividing

Table 1.1: Mechanisms of plasmid stability from [1]

Resistance to environmental selection
Complementation of chromosomal mutation
Post-segregational killing of plasmid free cells
Influences of plasmid size and form
Active plasmid partitioning
High copy number
Lowering the difference in specific growth rates
Influences of cultivation conditions
Separation of plasmid free and plasmid bearing cells
Integration into the chromosome

cell. This passing on of genetic material is dependent on the spatial positioning of the copies of that genetic material within the dividing cell. As can be seen in Figure 1.3b, if copies of the genetic material are distributed across the entire cell, both daughter cells will inherit that material. If, however, there is an imbalance and the genetic material is localised to one side of the dividing cell, a daughter cell without that genetic material will be produced.

Most genes that one encodes on plasmids are non-essential and therefore potentially burdensome to their host as the plasmid 'hijacks' the cell's machinery for their own purposes [56], though there are a number of exceptions to this rule [57]. This commonly leads to a lower growth rate in plasmid-bearing cells relative to their plasmid-free counterparts which in turn leads to the dilution and eventual removal of plasmid-bearing cells from the environment.

A number of mechanisms for improved plasmid stability have been detailed [1]. Table 1.1 lists these mechanisms. In Section 1.3 I will discuss the mechanisms that are available to try and prevent plasmid-free cells arising in this way. In Section 1.4 I will discuss methods for preventing plasmid-free cells from overtaking the population if they do arise.

1.3 Segregational Plasmid Stability

The faithfulness with which a plasmid is able to guarantee its inheritance by both daughter cells upon the division of the parent is its segregational stability. Two

evolved mechanisms for improving segregational stability are active partitioning mechanisms and high plasmid copy number.

1.3.1 Active Partitioning

There are three forms of active partitioning system found in bacterial plasmids: the type I *parABC*-like systems, the type II *parMRC*-like systems and the type III *tubRZ*-like systems [58]. The type I and II systems are more common and both are three component systems involving two proteins and a binding region on the plasmid. As an example of the efficacy of such mechanisms, the type I *par* mechanism on the P1 plasmid is effective enough that with a plasmid copy number of $3-4$ there is a loss rate of roughly one in 10^4 divisions [59].

1.3.1.1 Type I - *parABC*

In the *parABC* three component system the *parC* region on the plasmid can be bound to by the *cis*-encoded ParB proteins. This complex can then move within the cell, dependent on the distribution of ParA proteins. ParA is a *cis*-encoded ATPase which, when in a dimer with ATP, binds to the surface of the bacterial nucleoid, forming filaments with other bound ParA-ATP dimers [60]. The ParB of the ParB-plasmid complex binds to the ParA-ATP dimer, which promotes the conversion of ParA-ATP to ParA-ADP. This causes the ParA to unbind from the nucleoid.

It had been thought that ParB was immobilised by ParA and moved randomly once the ParA had become unbound. The movement of the ParA on and off the nucleoid is dependent on the position of the ParB-plasmids, leading to oscillations around regularly spaced equilibrium positions. However, it is possible that the disassembly of the ParA-ATP pulls the ParB along the ParA filament [60]. This allows the ParB to interact with the next ParA-ATP in the chain rather than randomly diffusing until it is immobilised by the next ParA that it bumps in to. A schematic of the model is shown in Figure 1.4. Computational modelling shows that this model produces a regular spacing of plasmids across the nucleoid of the cell, which fits experimental data from *Escherichia coli* plasmid pB171 [61]. The regular spacing of the plasmids ensures that there is equal distribution between two daughter cells

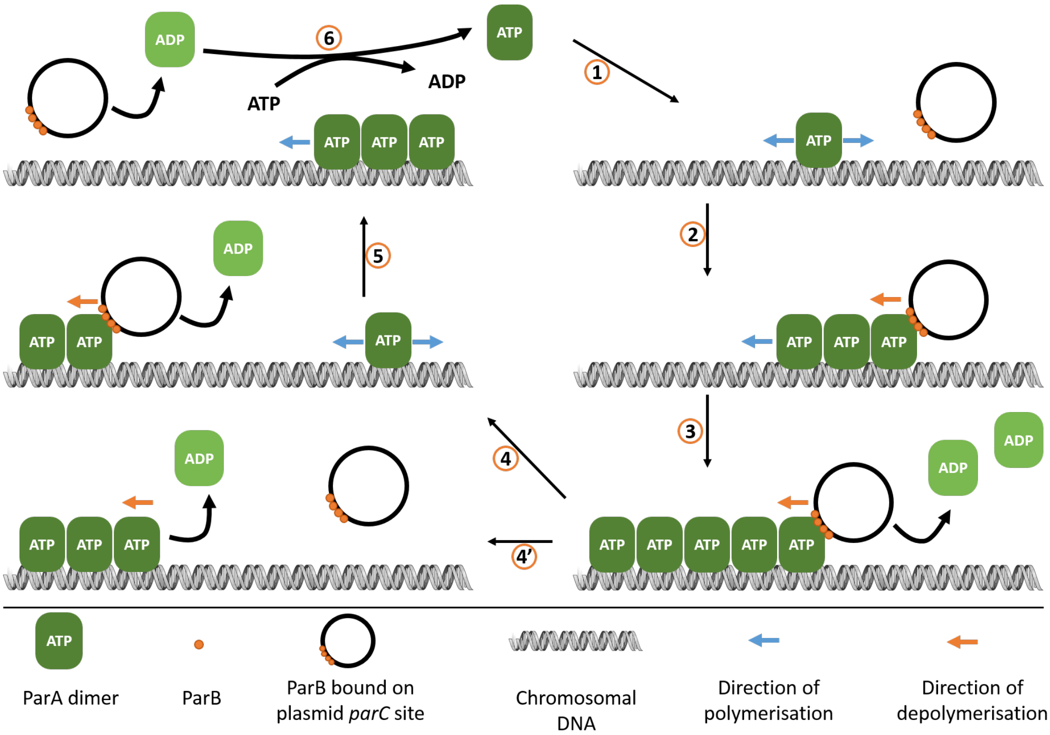

Figure 1.4: A model of plasmid positioning by *parABC*. (1) ParA-ATP dimers bind to the bacterial nucleoid. (2) ParB, bound to the *parC* region on the plasmid, associates with the nucleoid bound ParA-ATP. (3) The ParB promotes the conversion of ParA-ATP to ParA-ADP, leading to the unbinding of the ParA-ADP from the nucleoid. (4’) In the diffusion-immobilisation model, the ParB is then released from the unbound ParA-ADP and the plasmid diffuses randomly. (4) In the model theorised by [60] the ParB is dragged to the next bound ParA-ATP in the filament. (5) When the ParB reaches the end of a filament it randomly diffuses until it associates again with a ParA-ATP dimer. (6) The unbound ParA-ADP diffuses randomly and after a period of time until it reforms ParA-ATP. Adapted from [60]

at division.

1.3.1.2 Type II - *parMRC*

The model type II *par* system is the *parMRC* mechanism found in the *E. coli* R1 plasmid [62]. *parMRC* is a three component system with similar constituents to *parABC* [63]. The *parC* is a DNA sequence on the plasmid on to which the ParR adapter proteins bind. The ParM proteins bind with ATP and begin to form polar filaments, though these are disassembled upon hydrolysis of the ATP [64]. The ParR*C* complex binds to the ParM filament, stabilising the end [65]. Two ParM filaments can associate to form bipolar filaments with a ParR*C* complex at either

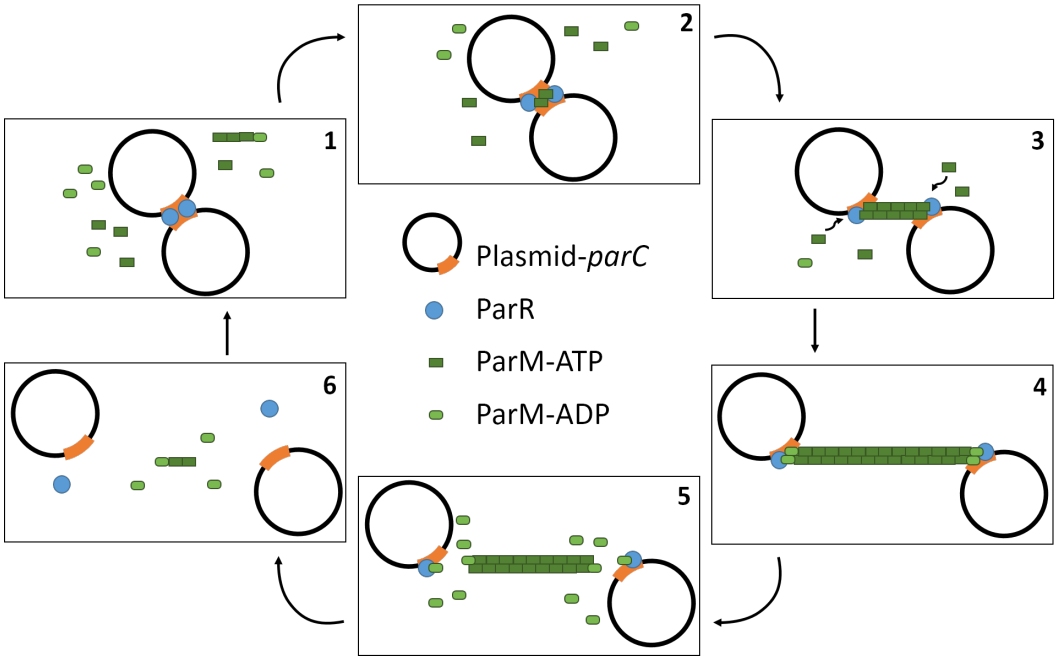

Figure 1.5: A model of plasmid positioning by *parMRC*. (1) ParR binds to the *parC* region of the plasmid. When two plasmids encounter each other they associate for a period of time. (2) Each ParR-*parC* complex is able to clamp a ParM-ATP subunit. (3) The ParM filament polymerises bidirectionally. (4) The filament continues to grow until the plasmids have been pushed to opposite poles of the cell. (5) Finally, the filament disassembles, leaving the plasmids at the poles for a period of time before they begin to diffuse randomly. Adapted from [66]

end. The binding of ParR*C* to the end of a ParM filament accelerates the filament's growth [62]. This leads to the bipolar filaments pushing the plasmids at either end apart and towards the poles of the host cell [58]. A model of the action of the *parMRC* system of the R1 plasmid is shown in Figure 1.5. Thus, whereas the type I systems used filaments to "pull" plasmids around a cell, the type II system uses filaments to "push" plasmids away from each other.

1.3.2 Plasmid Copy Number

If no active partitioning mechanisms are present, plasmids are assumed to be free to diffuse randomly throughout the cell. If it is assumed that cells divide perfectly in half, there is a 50% chance of each plasmid being passed on to a specific daughter cell. Then the distribution of plasmids in each daughter cell after division follows a Binomial distribution with probability of 0.5 and the number of trials equal to the number of plasmids, n. The probability that no plasmid will end up in a specific

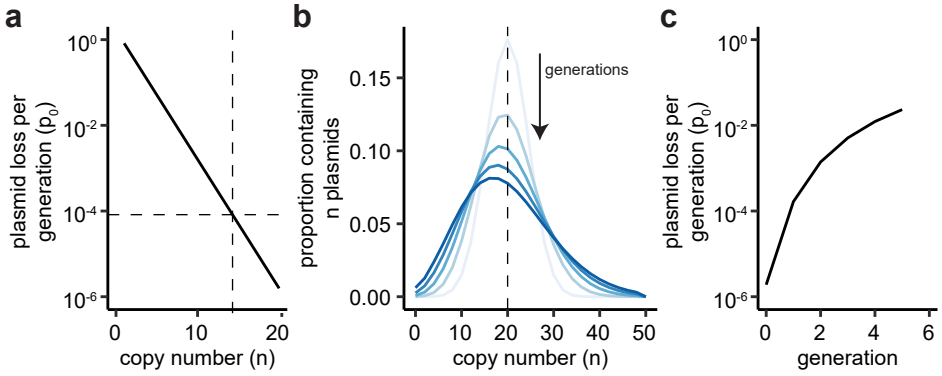

Figure 1.6: (a) The probability of producing a plasmid-free daughter cell based on the assumption of a random distribution of monomeric plasmids within the cell. The dashed lines show the copy number needed to replicate the stability of the partitioning mechanism found on the P1 plasmid. (b) With no copy number control the binomial distribution of plasmids between daughter cells at division leads to an increasing plasmid copy number variance as generations progress. In this model plasmids are binomially distributed between daughter cells and then replicate once before the daughter cells divide. The dashed line shows the initial plasmid copy number. (c) This variance in plasmid copy number leads to increases in the plasmid loss probability as the subset of the population with lower plasmid copy number increases.

daughter cell reduces to 0.5^n and since there are two daughter cells, the probability of producing a plasmid-free daughter cell after division is given by $p_0 = 2(0.5)^n = 2^{(1-n)}$ [29]. In order to achieve the the 10^{-4} loss rate attained by the P1 plasmid *par* mechanism a randomly distributed plasmid would need to have a copy number of just over 14, as can be seen in Figure 1.6a. As such, another mechanism for increasing plasmid persistence is to increase plasmid copy number, which is the approach often taken in industrial biotechnology. However, this model relies on a number of assumptions detailed below.

1.3.2.1 Copy Number Control

Variance in plasmid copy number at the point of cell division affects the rate at which plasmid-free cells arise [29]. Figure 1.6b shows how variance in plasmid copy number across a population increases when plasmids are binomially distributed to daughter cells on division. As the number of division cycles progress, the copy number distribution spreads leading to an increase in cells containing a low copy number of plasmids. As demonstrated in Figure 1.6a, as copy number goes

down plasmid loss probability exponentially increases. This leads to the increase in plasmid loss probability as the generations go by shown in Figure 1.6c. Thus, there must exist mechanisms that allow plasmids to attain a target copy number before cell division occurs again.

The pSC101 plasmid produces a protein, RepA, which binds to a site on the plasmid itself [67]. When the plasmid copy number is too low there is a deficiency in binding sites, leading to and excess of free RepA. The free RepA induces an SOS response in the host *E. coli* which delays cellular division. This enables the plasmids to continue to replicate until there are enough to deplete the free RepA concentration.

Replication of the plasmid ColE1 is primed by the transcription of the replication primer precursor, RNA II [68]. During its transcription the antisense RNA, RNA I, can form a kissing complex with RNA II, in which the unpaired bases in the hairpin loop of RNA I pair with those in the RNA II. This can then lead to an RNA I:RNA II duplex which causes the inhibition of primer formation and the abandonment of replication. The kissing complex is stabilised by the plasmid encoded Rom protein, which increases the likelihood of duplex formation. Using a mathematical model it was shown that there are optimal rates of transcription and degradation of RNA I, RNA II and the Rom protein, that strengthens their association, in order to achieve segregational stability with minimal metabolic burden [68].

The RepA mechanism in the pSC101 plasmid places a lower bound on the copy number by preventing cell division until a threshold copy number has been reached. Conversely, the replication control mechanism of the ColE1 plasmid places an upper bound on the copy number by preventing replication based on the concentration of RNA I and Rom in the cell.

Multimer Resolution The formation of plasmid multimers by homologous recombination results in an effective reduction in copy number [69]. As described above, a reduction in copy number results in a decrease in plasmid stability. The problem is exacerbated by the fact that a plasmid dimer, although counting as a single plasmid for the sake of segregation, contains two origins of replication. As such, a dimer is

twice as likely to replicate as a monomer leading to the accumulation of dimers and larger oligomers; the 'dimer catastrophe' [69].

In order to protect from the destabilisation affects of oligomer formation, the ColE1 plasmid has a mechanism for resolving those oligomers to monomers [29]. It involves the recruitment of host encoded proteins XerC, XerD, ArgR and PepA, to the *cer* site on the plasmid [70, 71].

A promoter in the *cer* multimer resolution site of the ColE1 plasmid produces an RNA called the regulator of cell division (Rcd) [72]. Over expression of this transcript was shown to delay cell division as long as tryptophanase was present. Tryptophanase catabolises tryptophan to pyruvate and indole which suggests that indole, which is usually associated with quorum sensing behaviour, acts as the signalling molecule to prevent division [71]. Stochastic simulations show that this delay may allow the plasmid dimers to be resolved to monomers. However, the model seems to suggest that the main benefit comes from a reduction in metabolic load from the dimers than from enhancing plasmid stability [73].

Plasmid Clustering An important assumption in the binomial distribution model is that the plasmids are positioned randomly within the cell. Since high copy plasmids, such as ColE1, do not appear to contain a *par* mechanism this would seem to be a reasonable assumption. However, localisation of ColE1 to the poles of their host cells has been observed [74]. The polar positioning of ColE1 was able to be displaced by the introduction of plasmid pBAD33, which occupied the polar positions instead. This led to the random positioning of the ColE1 plasmid and a decrease in segregational stability. Such clustering could be explained by nuclear exclusion combined with random diffusion but suggests that it is more likely that a regulated process is at work [66].

RK2 plasmids have also been shown to cluster in *E. coli* [75]. It was observed that the number of foci increased with cell length and growth rate, suggesting co-ordination with the host's cell cycle. The foci were regularly spaced as has been seen with the *parABC* system present. This may lead towards confirmation of the speculation that the origin of replication may have a role in recruiting the host cells

partitioning factors [66].

1.4 Plasmid-free Cell Viability

Certain molecules called bactericides have the ability to kill bacteria. Other molecules, called bacteriostats, can prevent growth and reproduction. They can be externally introduced to the environment, as with antibiotics, or produced by the cells themselves.

1.4.1 Environmental Selection

Environmental selection, in the form of antibiotics, is the most common form of ensuring plasmid maintenance currently used in synthetic biology. Antibiotics are added to the environment, complimented by an antibiotic resistance gene on the desired plasmid. A cell's survival is then only possible when carrying the plasmid bearing the resistance gene. The cost of removing and deactivating antibiotics from industrial fermentors means that they are not used [33]. Antibiotics are also unsuitable for clinical applications for a number of reasons including horizontal gene transfer of resistance genes [34, 35] and the disruption of the native microbiota [36].

1.4.2 Post-segregational Killing

In a similar manner to environmental selection mechanisms, post-segregational killing (PSK) doesn't prevent plasmid loss but instead prevents cells that have lost a plasmid from proliferating. The difference with this system is that the plasmid plays an active role in killing the daughter cells as well as keeping them alive.

Toxin-antitoxin (TA) systems are a type of post-segregational killing mechanism that work through the production of a long lasting toxin and its shorter lived antitoxin [76, 77], both encoded on the plasmid. While the plasmid is present in the cell, antitoxin is being produced in excess to neutralise the toxin. If, however, a plasmid-free daughter cell arises the antitoxin quickly degrades and is unable to be replaced due to lack of the necessary genes. The negative effects of the toxin are no longer prevented, leading to the killing or growth prevention of the cell.

The standard model for the function of TA systems is that a toxin, produced by the plasmid, has its toxic effect nullified by an antitoxin, also produced by the

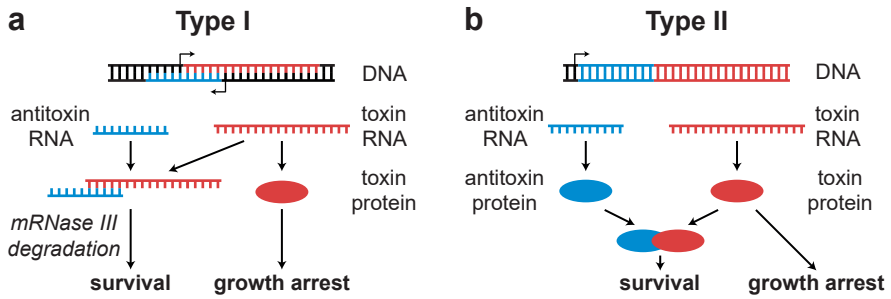

Figure 1.7: Mechanism of activity of type I and type II toxin-antitoxin systems.

plasmid. The toxin has a greater stability than the antitoxin and so persists in cells and daughter cells over a longer period. In a plasmid-bearing cell, the production of both toxin and antitoxin are balanced so that the deleterious effects of the toxin are not realised. If, on division, a plasmid-free daughter cell is produced, it will initially contain balanced levels of toxin and antitoxin but as the antitoxin degrades faster, an imbalance will develop and the toxin will have an effect.

Five types of TA system have been identified, though only one case of each type 4 and type 5 have been detailed [78]. They differ in regards to whether the toxin and antitoxin function post-transcriptionally or post-translationally, Figure 1.7. Further, the effects of the toxin can be either bactericidal or bacteriostatic [79]. The modes of action of the toxins differ and include the cleavage of mRNAs [80], inhibition of cell wall synthesis [81] and the inhibition of DNA replication [82]. TA systems have also been found in bacterial genomes, where it has been suggested they play a role in biofilm and persister cell formation [83].

In Chapter 5 I demonstrate the use of two different TA systems, hok/sok and axe/txe.

1.4.3 Bacteriocins

Bacteriocins are peptides and proteins that have a bactericidal affect on either a narrow or broad spectrum of other bacteria lacking resistance to the mechanism [84]. A common way of classifying bacteriocins is based on the size of the active protein, with subcategories that take into account other properties such as their biochemistry [85].

The modes of action of bacteriocins are varied, with many interacting with the cell wall to cause depolarisation or prevent synthesis [86]. Most groups of bacteriocin require extra enzymes to produce post-translational modifications, though a group of class II bacteriocins remain unmodified.

A number of examples of use of bacteriocin mechanism in engineered bacteria exist. *Lactococcus lactis* was engineered to release specific peptides when they detected the cCF10 pheromone of *E. faecalis* [87]. It was shown that this was effective in suppressing the *E. faecalis* population, even where commonly used antibiotics were ineffective. In a study that went one step further, a novel bacteriocin was produced from two others to create a bacteriocin that would only target *Pseudomonas aeruginosa* [88]. The system detected a quorum sensing molecule secreted by the *P. aeruginosa* and it was able to show inhibition of the growth of the pathogenic bacteria.

Bacteriocins have also been used explicitly as a plasmid stability mechanism [89]. The Lcn972 bacteriocin was able to stabilise previously unstable plasmids in *L. lactis*.

1.5 Metabolic Burden

Plasmid encoded genes imposes a burden on the host due its requirements for replication as well as transcription and translation. Although plasmids tend to add a minimal extra amount of DNA to a cell, evidence exists that there is evolutionary pressure on them to reduce the cost of carrying and replicating the extra DNA. The AT richness of plasmids compared to the average host chromosome suggests that they are optimising for reduced replication burden, as those are the less energy expensive base pairs to produce [45]. Although larger plasmids produce a greater impact on cell growth, ATP consumption for plasmid gene expression is a far greater source of burden [57].

A recent advance in the prediction of the burden that a given plasmid will place on a host uses a cell-free protocol [90]. By determining the reduction in the fluorescence of a control genetic circuit, the burden that a competing plasmid will

place on a host bacteria can be calculated. This system, along with an earlier one from the same group [91], can be used to design plasmids with lower burden.

An interesting example of a naturally occurring mechanism for the reduction of host metabolic burden at the cost of, what one would have assumed would prove to be, a plasmid fitness advantage is that of the fertility inhibition (*fin*) systems. Conjugative plasmids encode the *tra* gene which produces the machinery for horizontal gene transfer. However, the conjugative efficiency of the R1 plasmid is reduced 1000-fold by the *fin* gene [92]. By blocking the conjugative machinery, the *fin* system reduces the metabolic burden on the host as well as preventing the increased susceptibility to bacteriophages caused by expression of *tra* genes [92]. Although increased conjugation allows plasmids to more effectively invade and persist in host populations, the increased growth rate and vertical transmission of *fin* expressing bacteria leads to a competitive advantage [93].

1.5.1 Host/plasmid co-evolution

An assumption that is often made, and that has been presented here, is that by placing plasmids into a host a reduction in the host growth rate, however small, is inevitable. However, both the host and the plasmid are evolvable and, as such, if there is a selective pressure towards increased growth rate the two could evolve together to increase the growth rate.

Such a result was shown after the selective growth of a strain of *E. coli* with a plasmid for 500 generations [94]. It was observed that, as expected, when in competition with plasmid-free bacteria, plasmid-bearing bacteria had a relatively lower fitness. However, after 500 generations of forced co-existence, the plasmid-bearing bacteria out-competed their plasmid-free counterparts. Further, it was shown that even with the original unevolved plasmid, the same fitness advantages were seen, indicating that it was the bacterial genome that altered rather than that of the plasmid. A later study was able to show similar results [92].

Neither of these studies directly measured growth rates, rather competition against plasmid-free cells was used. Since growth rates are often used as a standard measure of fitness it is difficult to extrapolate these results.

1.6 Synthetic Microbial Communities

As the complexity of synthetic biological circuits increases, the burden that they place on their hosts likely does as well. In addition, as the range of biological products that we wish to produce expands, there is a likelihood that the target product or by-products in the chain will prove toxic to one of the standard organisms that we use as hosts. These problems prompt the use of communities of bacteria to carry out separate functions while interacting.

1.6.1 Coexistence of Microbial Communities

Interactions between species comprising a community can be positive or negative, and the combinations of these interactions can be used to classify the community relationships, as in Figure 1.8. Examples of positive interactions include the secretion of a metabolite that another species is unable to produce for themselves [95] or the degradation of environmental toxins [96]. Negative interactions include the secretion of antimicrobials into the environment [97] or the use of a resource needed by a competitor [98].

In many applications of microbial communities, even if the constituent species are not competing for a substrate, space is often a limiting factor over which they must compete. This leads to a problem, commonly known as the competitive exclusion principle, in which two species competing for a single resource cannot stably coexist. This results in many current systems having to use physical separation of the constituent strains to maintain multispecies communities [99]. Synthetic biology is now being used to produce communities that can stably coexist through the addition of extra interactions that balance the existing competitive relationships.

Communication is a key tool in organising these new interactions. A predator-prey system has been developed that uses quorum molecules to provide direct interaction between a “predator” bacteria and a “prey” bacteria [100]. The quorum molecule produced by the prey strain represses toxin production in the predator strain, thus rescuing it. The quorum molecule produced by the predator strain induces toxin production in the prey, thus killing it. This leads to the characteristic population oscillations of a predator-prey system [101].

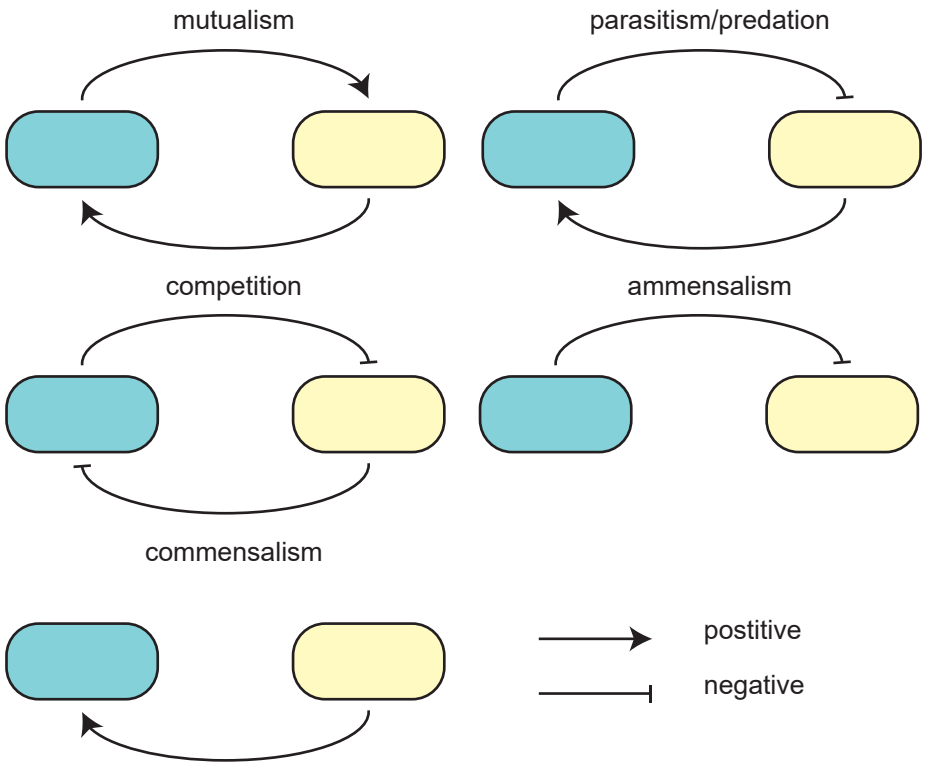

Figure 1.8: Types of community relationship.

More recently, two self-limiting strains of bacteria have been shown to be able to co-exist [102]. In this system, each strain produces and senses its own quorum molecule. This is used to trigger the production of a lysis gene when the concentration of the quorum molecule, and therefore of the producing strain, is high. The density linked lysis leads to oscillations in each population which free up space and substrate for the other population.

1.6.2 Industrial Applications of Microbial Communities

Production of biofuels from cellulosic mass has been one of the main applications for the development of microbial communities. In traditional fermentation to produce bioethanol, the cellulosic material must first be hydrolysed to release the sugars [103]. This is often done chemically but releases molecules which inhibit the later fermentation. Further, many organisms can't efficiently ferment all of the sugars that are found in most cellulosic materials. A community of two *E. coli* strains has been engineered to carry out the hydrolysis and fermentation of the hemicellulose

xylan [99]. One of the strains produces the hemicellulases to break down the xylan, while the other strain ferments the sugars that are released. A similar study used a community of fungi and bacteria to produce isobutanol [104].

1.7 Modelling of Dynamical Biological Systems

Mathematical models provide an exact form for describing ones assumptions regarding a system. Detailed descriptions of molecular, atomic and even sub-atomic interactions can be written mathematically. However, creating models with such levels of detail are often not best for making useful predictions. A quote attributed to John von Neumann suggests why; "With four parameters I can fit an elephant, and with five I can make him wiggle his trunk." Complex models with lots of unknown parameters can be made to fit any dataset. As such, we must understand the level of abstraction at which our models can still produce useful predictions but that provide enough simplicity that their misuse is restricted.

There are two main forms of modelling dynamical biological systems: deterministic and stochastic. Deterministic modelling uses systems of ordinary differential equations (ODEs) to describe how species within the model change over time. One must assume that the species within the system can be modelled continuously and as such can be described in terms of concentration rather than absolute number. When the number of a species is particularly small, this assumption does not hold as stochastic effects can dominate. Further, one must make the assumption that the species within the system are well-mixed, so species interact immediately and equally. Partial-differential equations can be used if this assumption does not hold. On the scale of bacterial cultures, where there are often $> 10^8$ cells, deterministic modelling is usually appropriate.

Conventional wisdom suggests that biological systems are deterministic if one has complete knowledge of the state of the system. However, the requirement of complete knowledge is presently unattainable. Further, the computational power necessary to simulate such a system is also not currently available. As such, the models that we build necessarily simplify the world. The effects of those small

details that have been removed can still be seen in the experimental data. Due to the law of large numbers, when the number of interacting species is large, a deterministic approximation is appropriate, as stated above. Stochastic modelling can be used in domains where species numbers are low. The randomness of interactions is modelled using a joint probability distribution for a particular reaction to happen at a given time.

1.7.1 Stochastic Simulation

Simulating stochastic systems can be computationally intensive. The commonly used Gillespie algorithm works by simulating every single reaction that takes place. While this provides an exact simulation of the system, if there are a mixture of fast and slow reactions, it can take an inordinate amount of time to run. Approximate methods exist which determine how many reactions can be simulated at once before the probabilities of each reaction occurring change "too much".

In the process of carrying out the work for this thesis one such approximate method, Tau-leaping, was implemented to run simulations in parallel on graphics processors (GPGPUs). Although this program was not in the end used for any of the simulations presented in this thesis, the code is available online.

1.7.2 Promoter Activity

It is often necessary to model the activities of promoters used in synthetic biological systems. The activities, often measured through a fluorescent reporter, are dependent on the concentration of an inducer or repressor. Three main paradigms exist to model these activities when gene expression is assumed to be at steady state.

Since promoter activation curves tend to be sigmoid in form, a logistic function is sometimes fit to experimental data. The concentration of inducer at which the promoter is half-maximally activated, the maximum promoter activity and the steepness of the response can fitted from experimental data. This model is purely descriptive.

The second model, the Hill function, has slightly more grounding in the mechanism of promoter activity. The function was developed to describe the dynamics of

ligands binding to macromolecules but in the case of promoters describes inducing molecules binding to promoters. Once again, the concentration of inducer at which the promoter is half-maximally activated and the maximum promoter activity are present. Here, however, the steepness of the curve is defined by a cooperativity parameter which conceptually describes how the binding of the first ligand affects the binding of subsequent ligands i.e. does it increase, decrease or have no effect on the rate.

Finally, the Shea-Ackers formalism produces models of promoter activity based on statistical thermodynamics [105]. The binding states of transcription factors and RNA polymerase to the promoter are defined and given weights [106]. The probability of the transcription from the promoter is then given by the ratio of states that can produce a transcript over all possible states. This formalism is the most mechanistic of the three described and explicitly includes parameters such as promoter leakiness.

1.8 Bayesian Parameter Inference

Once a model has been constructed, the parameters within the model need to be given values. This can be done by searching the literature, though experimental differences can lead to different measurements for the same parameter. For example, the growth rate of *E. coli* varies drastically depending on temperature, shaking, carbon source etc. As such, unless the conditions under which literature parameters were ascertained can be controlled for, they should only be used as a guide. Further, some model parameters may be impractical to directly measure experimentally.

Many optimisation methods, which infer model parameters by fitting the model to experimental data, have been developed. All of these methods work by minimising a cost function. This is a measure of how closely the model fits the experimental data given a set a parameters. Various forms of regression analysis enable this form of optimisation but they have limitations, such as requiring all parameters to be linearly independent. In many mechanistic models of biological systems, this assumption can't be met. As such, in systems in which there is correlation between

parameters, the point estimates of those parameters can be extremely dependent on the initial conditions of the regression.

A Bayesian approach allows for prior knowledge, such as parameters from the literature, to be used to inform where to look for the "true" parameter values when fitting to experimental data [107]. In addition, the Bayesian approach informs us of the uncertainty in the parameter estimates by producing posterior distributions rather than point estimates. This is formally written as:

$$p(\theta|D) \propto L(\theta|D)p(\theta) \tag{1.1}$$

where $p(\theta|D)$ is a posterior distribution for the parameters θ given some data D, $L(\theta|D)$ is the likelihood function and $p(\theta)$ incorporates our prior information. How one should choose prior distributions is contentious but a reasonable recommendation is that they should be based on the researcher's prior knowledge of the system [108], which often includes previous experimental results and values from the literature.

It is often not possible to analytically derive an expression for the posterior distribution. As such, Monte Carlo methods, such as Markov chain Monte Carlo (MCMC), can be used to produce a sample from the posterior [109].

1.8.1 Approximate Bayesian Computation

For some models it can be impossible to write an expression for the likelihood. In these situations approximate Bayesian computation (ABC) can be used to simulate data from the model with which to provide a comparison with experimental data [110].

The ABC rejection sampler is the simplest form of ABC algorithm. The algorithm is defined as follows

where D is the experimental dataset, f produces a simulated dataset, D^*, given a set of parameters, d is the distance function, and ε is a tolerance level. As the threshold tends to zero, the accepted parameter vectors are a sample from the true posterior

Algorithm 1 ABC rejection sampler

for $i = 1$ to N **do**
  repeat
    Sample θ^* from the prior distribution $\pi(\theta)$
    Simulate D^* from the model $f(\theta^*)$
  until $d\{D^*, D\} \leq \varepsilon$
  set $\theta_i = \theta^*$
end for

distribution, $p(\theta|D)$ [108]. However, just setting the threshold to zero would lead to too many rejections when sampling the whole prior space. This is a weakness of the ABC rejection sampler that is addressed with ABC-SMC (sequential Monte Carlo) [111].

The ABC-SMC algorithm progresses through iterations, narrowing the threshold for acceptance and sampling new parameters based on previously accepted parameters. The algorithm is defined as follows

Algorithm 2 ABC-SMC algorithm

Initialise $\varepsilon_1 > \varepsilon_2 > ... > \varepsilon_T$
for $t = 0$ to T **do**
  for $i = 1$ to N **do**
    repeat
      if $t = 0$ **then**
        Sample θ^{**} from prior distribution $\pi(\theta)$
      else
        Sample θ^* from $\{\theta_{t-1}^{(i)}\}$ with weights w_{t-1}
        Generate θ^{**} by applying perturbation kernel $K_t(\theta|\theta^*)$
      end if
      if $\pi(\theta^{**} = 0)$ **then**
        Return to line 5
      end if
      Simulate D^* from the model $f(\theta^{**})$
    until $d\{D^*, D\} \leq \varepsilon_t$
    Set $\theta_{t-1}^{(i)} = \theta^{**}$
    Calculate weights $w_t^{(i)} = \begin{cases} 1 & \text{if } t = 0 \\ \frac{\pi(\theta_t^{(i)})}{\sum_{j=1}^{N} w_{t-1}^{(j)} K_t(\theta_{t-1}^{(j)}, \theta_t^{(i)})} & \text{if } t > 0 \end{cases}$
  end for
  Normalise the weights
end for

In a similar fashion to the ABC rejection sampler, the ABC-SMC algorithm rejects parameters if they produce a simulated data set that is too dissimilar from the experimental dataset. The initial parameters that are chosen for the simulation are based purely on the priors specified by the researcher. However, in following rounds, the new parameters are produced by sampling the parameters that were produced from the previous round, with a bias towards those that produced the best fits, and perturbing them.

1.9 Thesis Outline

This thesis will focus on the control of synthetic biological systems at the plasmid and population levels. It is organised as follows:

In Chapter 2 I detail the materials and methods used to produce and analyse the data in this work.

In Chapter 3 I describe a system that uses competitive inhibition and the dynamical control of an antimicrobial peptide to stabilise two competing bacterial populations. Using mathematical modelling I show how this system can be tuned to produce a range of behaviours with parameters that can be simply controlled experimentally.

In Chapter 4 I construct and characterise the population stabilisation system. Preliminary experiments in a bioreactor are undertaken and demonstrate the ability of the system to function as designed. Finally, a more complete mathematical model is constructed and, using Bayesian methods, an understanding is developed as to how to produce desired behaviours in future experiments.

In Chapter 5 I detail the examination of three systems for the enhancement of plasmid stability. A model is developed to probe these systems and Bayesian techniques are used to understand experimental data. The systems are tested using a therapeutically relevant strain of *E. coli in vitro* and in an *in vivo* mouse tumour model.

Finally, Chapter 6 provides a summary of this thesis.

1.10 Work Carried Out by Other Individuals

The work in Chapter 5 was carried out in collaboration with Tanel Ozdemir (Barnes Lab, UCL). He completed the cloning of the fluorescent reporter plasmids with the toxin-antitoxin systems and all of the *in vitro* plasmid loss experiments were undertaken together. Initial construction of the luminescent plasmids was undertaken by Tanel Ozdemir and myself, while the promoters were cloned by Anjali Doshi (Danino Lab, Columbia University, USA). Anjali also carried out all of the *in vivo* experiments.

The search for oscillations and equilibria in the SPoCK model, using approximate Bayesian computation, in Chapter 4 was carried out with Behzad Karkaria (Barnes Lab, UCL).

Mike Sulu (UCL) and Noelle Colant (UCL) aided me with the running of the chemostat experiment in Chapter 4.

Work carried out throughout my candidature has produced the following articles:

- Miriam Leon, Mae L Woods, Alex J H Fedorec, and Chris P Barnes. A computational method for the investigation of multistable systems and its application to genetic switches. *BMC Systems Biology*, 10(1):130, 2016

- Alex J H Fedorec, Tanel Ozdemir, Anjali Doshi, Luca Rosa, Oscar Velazquez, Tal Danino, and Chris P Barnes. Two new plasmid post-segregational killing mechanisms for the implementation of synthetic gene networks in *E. coli* [Submitted]

- Tanel Ozdemir, Alex J H Fedorec, Tal Danino, and Chris P Barnes. Synthetic biology and engineered live biotherapeutics: Toward increasing system complexity. *Cell Systems*, 7(1):5–16, 2018

- Alex J H Fedorec, Behzad Karkaria, and Chris P Barnes. Dynamic stabilisation of multi-species bacterial consortia using bacteriocins and competitive exclusion [In preparation]

Chapter 2

Materials and Methods

2.1 General Methods

2.1.1 Plate Reader Growth Rate Assays

Strains were spread on selective LB agar plates from glycerol stocks using an inoculation loop. The plates were grown overnight at 37°C. Single colonies were picked and inoculated into 5mL of selective LB media in a 15mL Falcon tube. The tubes were incubated overnight at 37°C and 200 rpm in a 2.5 cm orbital shaking incubator. The cultures were diluted 1:1000 in to 10 mL of fresh LB media without antibiotics. Six 200 µL replicates of each strain were pipetted in to a black clear-bottom 96-well microtitre plate (Greiner Bio-one, Germany). Any empty wells were filled with LB media and the plate was sealed with a Breathe-Easy sealing membrane (Sigma-Aldrich, UK). The plate was immediately placed in a POLARstar Galaxy microplate reader (BMG LABTECH, UK) at 37°C with continuous 3mm double orbital shaking for 20 hours with absorbance readings at 540 nm taken every 10 minutes.

2.1.2 Flow Cytometry Assays

Flow cytometry was performed on an Attune NxT Acoustic Focusing Cytometer with Attune NxT Autosampler (Thermo Fisher Scientific, UK). The wells of a round-bottomed 96-well microtitre plate (Nunc, Denmark) were filled with 200 µL of PBS. 1 µL of sample was inoculated in to each well and the solution mixed by pipetting up and down. The plate was immediately transferred to the flow cy-

Table 2.1: Fluorescence channels for the Attune NxT flow cytometer.

Excitation laser	Emission filter (nm)	Fluorescent protein
Violet (405 nm)	440/50	BFP
	512/25	CFP
	603/48	
	710/50	
Blue (488 nm)	530/30	GFP and YFP
	590/40	dsRed
	695/40	
	780/60	
Yellow (561 nm)	585/16	dsRed
	620/15	mCherry
	695/40	
	780/60	

tometer for sampling. The Attune NxT Autosampler was set sample 20 µL from each well with 2 mixing cycles and 4 rinses between each sample. Forward and side scatter height and area measurements were always recorded. Height and area measurements in the appropriate fluorescent channels were also recorded. The fluorescence channels are detailed in Table 2.1 and the expected emissions spectra (http://www.spectra.arizona.edu) for each of the available lasers are shown in Figure 2.1.

2.1.3 Flow Cytometry Data Processing

In order to ensure the capture of all bacterial events in the flow cytometry sampling, a threshold was used that also captured a considerable amount of non-bacterial events. This can be problematic as the small size of bacteria leads to the bacterial population appearing very close to this debris. Further, more than one bacterium can be read as a single event due to clumping. A computational pipeline was developed to automate the gating of singlet bacteria while excluding background debris and bacterial doublets. This process is important to automate as it removes operator bias associated with the manual drawing of gates and speeds up the analysis when large numbers of samples are recorded.

The method fits two dimensional mixture models of one and two clusters to the forward-scatter-height and side-scatter-height variables with 95% confidence.

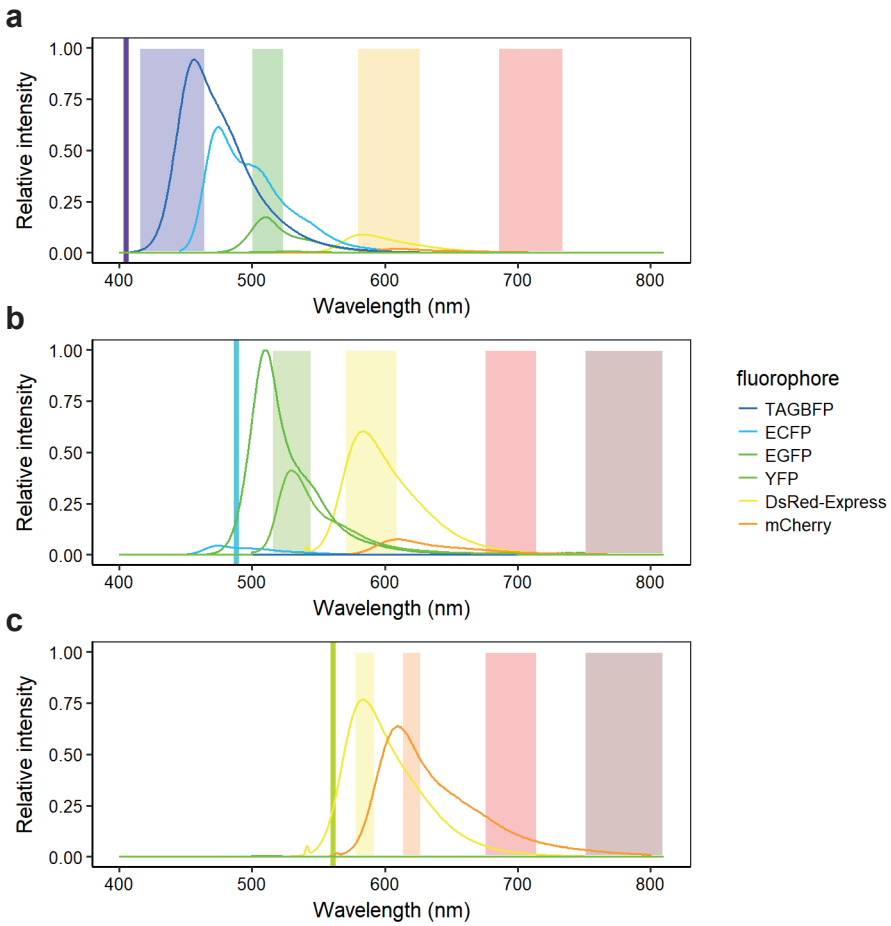

Figure 2.1: Fluorescence emissions spectra for available flow cytometry lasers.

Using integrated completed likelihood (ICL) [115] the number of clusters that best fits the data is determined. This allows for compatibility with experiments in which debris has been removed by hand by an experimenter or if not many debris events are recorded. For any clusters found, their median positions are determined and if a cluster is below a pre-determined threshold it is considered debris and removed.

Doublets are then removed in one of two ways. Initially, a linear model was fitted to the side-scatter-height versus side-scatter-area data and points falling too far away were removed as doublets. An update to the Attune NxT software implemented area scaling which is calibrated automatically during the start-up procedure. The area scaling allows the assumption that singlets will lie on, or close to, the line side-scatter-height = side-scatter-area. As such, for later experiments this method is used for doublet discrimination.

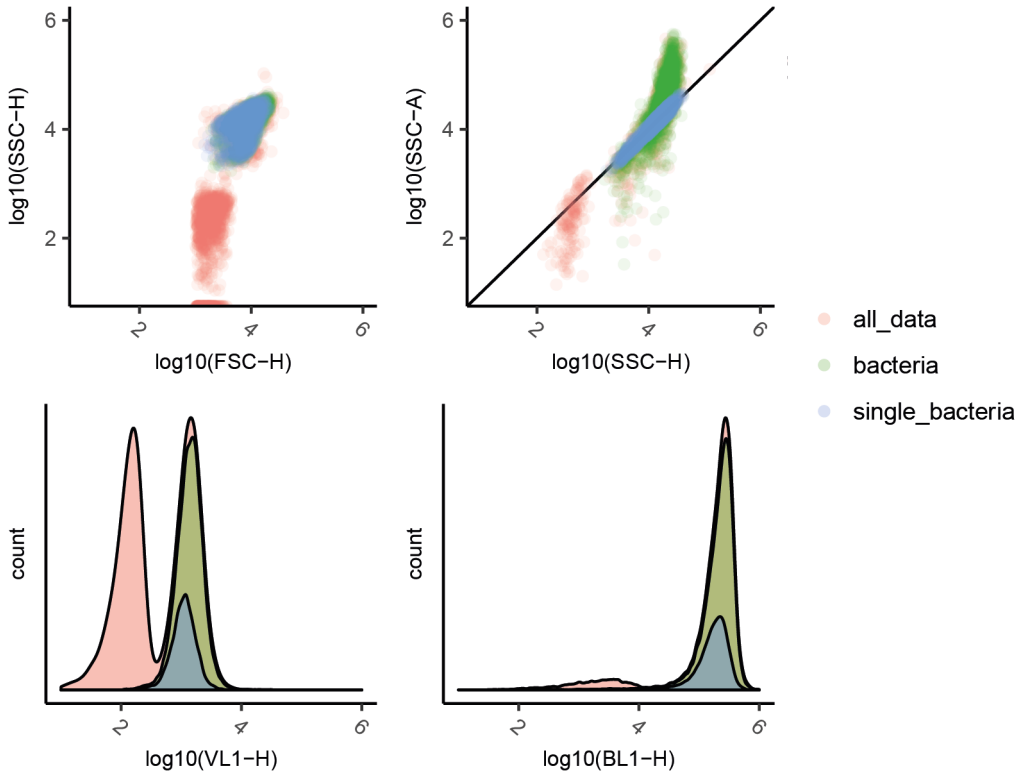

Figure 2.2: Processing of flow cytometry data to remove debris and doublets.

The output from this processing are standard flow cytometry .fcs files with debris and doublets removed.

2.2 Methods for SPoCK

A number of plasmids were created for the work in Chapter 4. These plasmids are listed in Table 2.2. All plasmid sequences were confirmed with Sanger sequencing. Where the created plasmids have used SEVA vectors, the SEVA naming convention for antibiotic resistance and origin of replication has been adhered to, though the "SEVA" prefix has been removed as the plasmids don't necessarily adhere to the SEVA standards. The SEVA convention for naming of the "cargo" region has not been adhered to as the extensibility of this naming convention is limited. The cloning procedures for the construction of some of these plasmids are detailed below.

Table 2.2: Plasmids used in this chapter.

Plasmid	Description	Source
pSEVA281	pUC ori, KnR, MCS	SEVA collection
pSEVA471	SC101 ori, SmR, MCS	SEVA collection
pSEVA637	BBR1 ori, GmR, GFP	SEVA collection
pSEVA191	BR322 ori, ApR, MCS	SEVA collection
pBAD-mTagBFP2	BR322 ori, ApR, araC ← ParaBAD → mTagBFP2	[116]
pTD103luxI_sfGFP	ColE1 ori, KnR, luxR ← PLux → sfGFP-LAA, luxR ← PLux → luxI-LAA	[117]
pKDL071	ColE1 ori, KnR, GFPmut3b ← PLtetO-1, lacI ← PLtetO-1, Ptrc → tetR, Ptrc → mCherry	[118]
pHK11	BR322 ori, ApR, mccV	[119]
pMPES:V	p15A ori, CmR, ProTeOn → cvaC + cvi, mccV*	[120]
pMPES_AF01	p15A ori, CmR, PLtetO-1 → GFPmut3b + cvaC + cvi, mccV*	this work
p28_AF02	pUC ori, KnR, luxR ← PLux → sfGFP-LAA + tetR	this work
p28_AF03	pUC ori, KnR, araC ← ParaBAD → mTagBFP2 + luxI-LAA	this work
p28_AF04	pUC ori, KnR, AF02, AF03	this work
p27_AF04	SC101 ori, KnR, AF02, AF03	this work
p23_AF04	BBR1 ori, KnR, AF02, AF03	this work
p29_AF04	BR322 ori, KnR, AF02, AF03	this work
p28_AF041	p28_AF04 with GFP removed	this work
p27_AF041	p27_AF04 with GFP removed	this work
p23_AF041	p23_AF04 with GFP removed	this work
p29_AF041	p29_AF04 with GFP removed	this work
pAF006	ColE1 ori, KnR, GFPmut3b ← PLtetO-1, mccV ← PLtetO-1, Ptrc → tetR, Ptrc → mCherry	this work
pAF0061	ColE1 ori, KnR, GFPmut3b ← PLtetO-1, mccV ← PLtetO-1, Ptrc → tetR	this work
pAF0062	ColE1 ori, KnR, GFPmut3b ← PLtetO-1, mccV ← PLtetO-1	this work
pTNS1	Tn7 transposase expression, R6K ori, ApR	[121]
pTn7-M-Pem7_mCherry	R6K ori, KnR, Pem7→mCherry + GmR transposable cassette	SEVA collection

Table 2.3: Bacterial strains used in this chapter.

Strain	Genotype	Source
E. coli NEB 5-alpha	*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 ϕ80Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	New England BioLabs
E. coli Top10F'	F'[lacI^q Tn10(tet^R)] *mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu)7697 galU galK rpsL endA1 nupG*	Invitrogen
E. coli MG1655	K-12 F$^-$ λ^- *ilvG$^-$ rfb-50 rph-1*	Prof. John Ward (UCL)
EcM-GmRed	*E. coli* MG1655 with chromosomal gentamicin resistance and mCherry	this work

Table 2.4: Primers used in this chapter. Underlined sequences show restriction sites added to the amplified fragments. Upper case letters show the annealing sequence.

Primer	Sequence
P.XmaI.luxR.F	cgatcacccgggCACCTCGAGTTAATTTTTAAAGTATGGGC
P.ApaI.GFP.R	cagctggggcccTAAGCTTTTACGCTGCAAGGGC
P.ApaI.tetR.F	gtgactgggcccCGGCGAAGCTAGGGAC
P.SbfI.tetR.R	gtcaatcctgcaggAGTCCGAGCTCTCAAGAC
P.luxI.F	AGGATCGTACAGGTTTACGCAAG
P.SpeI.luxI.R	ggctcgactagtTAGCACGCGTTTACGCTG
P.delGFP.F	GCATGGATGAGCTCTACAAAGC
P.delGFP.R	gcactagggcccATTCGACTATAACAAACCATTTTCTTGC
P.HindIII.mccV.F	tgacagaagcttTGCCCTTCCCTAGAGAATCC
P.BglII.cvaCi	gcatgcagatctGTGGTAATGGGATAGAAAGTAATGGG
P.PLtetO-GFP.SacI.F	atcgtagagctcCGTCTGTGCAAGTACTACTGT
P.PLtetO-GFP.XbaI.R	atgctatctagaAACTGGCACGACAGGTTTC
P.SEVA.ORI.F	CGGTGCTCAACGGGAATC
P.SEVA.ORI.R	TGTTGTTTGTCGGTGAACGC

The bacterial strains used in Chapter 4 are detailed in Table 2.3. Electrocompetent cells were produced using the mannitol-glycerol step protocol [122].

The primers used for PCR amplifications are detailed in Table 2.4. The annealing sequences were designed using Primer3Plus [123] and annealing temperatures were calculated using the NEB Tm calculator (http://tmcalculator.neb.com).

The analysis of the flow cytometry data followed the pipeline detailed in Section 2.1.3.

2.2.1 Cloning of AF02

The luxR ← PLux → sfGFP fragment was PCR amplified from the pTD103luxI_sfGFP plasmid using primers P.XmaI.luxR.F and P.ApaI.GFP.R with an annealing temperature of 67°C. The tetR fragment was PCR amplified from the pKDL071 plasmid using primers P.SbfI.tetR.R and P.ApaI.tetR.F with an annealing temperature of 66°C. The PCR mixes were run on a 0.7% agarose gel and the 1769 bp band was extracted for the PLux fragment and 688 bp band for the tetR fragment. The PLux fragment was digested with XmaI and ApaI, the tetR fragment was digested with SbfI and ApaI and the destination vector pSEVA281 was digested with XmaI and SbfI. The digestions were cleaned using a PCR cleanup kit. The two fragments and the vector were ligated together to form the plasmid p28_AF02 shown in Figure 2.3, and transformed into *E. coli* DH5α by electroporation.

2.2.2 Cloning of AF03

The luxI fragment was PCR amplified from the pTD103luxI_sfGFP plasmid using primers P.luxI.F and P.SpeI.LuxI.R with an annealing temperature of 68°C. The PCR mix was run on a 0.7% agarose gel and the 711 bp band was extracted for the luxI fragment. The luxI fragment was digested with EcoRI and SpeI and the destination vector pSEVA281 was digested with SphI and SpeI. The digestions were cleaned using a PCR cleanup kit. The pBAD-mTagBFP2 plasmid was digested with SphI and EcoRI and the araC ← ParaBAD → mTagBFP2 fragment at 2050 bp was gel extracted. The two fragments and the vector were ligated together to form the plasmid p28_AF03 shown in Figure 2.4, and transformed into *E. coli* DH5α by electroporation.

2.2.3 Cloning of AF04

The p28_AF02 plasmid was digested with XmaI, SbfI and ApaLI (in order to cut the unwanted fragment). The 2424 bp AF02 fragment, consisting of luxR ← PLux → sfGFP tetR, was gel extracted. The p28_AF03 plasmid was digested with XmaI and SbfI and cleaned using a PCR cleanup kit. The AF02 fragment was ligated into the p28_AF03 plasmid, to form plasmid p28_AF04 shown in Figure 2.5, and

transformed into *E. coli* DH5α by electroporation.

Versions with different origins of replication were created by PCR amplification of the relevant SEVA plasmid with primers P.SEVA.ORI.F and P.SEVA.ORI.R. The PCR mixes were run on 0.7% agarose gels and the bands at the expected locations were extracted. The fragments and plasmid p28_AF04 were digested with AscI and FseI and cleaned with a PCR cleanup kit. The plasmid p28_AF04 was digested with AscI and FseI and the band at 6654 bp was extracted. The respective origin of replication fragments and the digested plasmid were ligated together and transformed into *E. coli* DH5α by electroporation.

2.2.4 Cloning of AF041

The plasmid p28_AF04 was PCR amplified using primers P.delGFP.F and P.delGFP.R with an annealing temperature of 65°C. The PCR mix was run on a 0.7% agarose gel and the band at 6871 bp was extracted. The fragment was digested with ApaI and cleaned with a PCR cleanup kit. It was then ligated to create the re-circularised p28_AF041 plasmid, shown in Figure 2.6, and transformed into *E. coli* DH5α by electroporation.

Versions with different origins of replication were created in the same way as for AF04.

2.2.5 Cloning of pAF006, pAF0061 & pAF0062

The mccV cassette from plasmid pHK11 was PCR amplified with primers P.HindIII.mccV.F and P.BglII.cvaCi with an annealing temperature of 67°C. The PCR mix was then cleaned using a PCR cleanup kit, digested with HindIII and BglII and cleaned again. Plasmid pKDL071 was digested with HindIII and BamHI, run on a 0.7% agarose gel and the band at 5795 bp was extracted. The mccV fragment was ligated into the digested pKDL071 vector to create plasmid pAF006, shown in Figure 2.7, and transformed into *E. coli* DH5α by electroporation.

pAF0061 was created by digestion of pAF006 with XhoI and SacI, which was then run on a 0.7% agarose gel and the band at 8262 bp was extracted. The fragment was then end-blunted using T4 DNA polymerase, after which the fragment was

cleaned using a PCR cleanup kit. The fragment was then ligated with electroligase and transformed into *E. coli* Top10F' by electroporation.

pAF0062 was created in the same way but the original digestion was with XhoI and NcoI to produce a fragment of 7523 bp.

2.2.6 Cloning of pMPES_AF01

The PLtetO-1 promoter and GFPmut3b fragment was PCR amplified from plasmid pKDL071 with primers P.PLtetO-GFP.SacI.F and P.PLtetO-GFP.XbaI.R with an annealing temperature of 66°C. The PCR mix was run on a 0.7% agarose gel and the band at 980 bp was extracted. The fragment was digested with XbaI and SacI and cleaned using a PCR cleanup kit. Plasmid pMPES:V was digested with XbaI, SacI, and BamHI. It was digested into three parts to avoid loss of DNA when cleaning large fragments in Qiagen QIAquick columns. The digested plasmid was run on a 0.7% agarose gel and the bands at 3.9 Kbp and 9.4 Kbp were extracted. The GFP fragment and the two vector fragments were ligated together using electroligase and transformed into *E. coli* Top10F' by electroporation.

2.2.7 Insertion of GmR and mCherry into *E. coli* MG1655 Genome

A gentamicin resistance gene and mCherry driven by the Pem7 promoter were inserted into the genome of *E. coli* MG1655 to produce a selectable and detectable strain. This was done using the mini-Tn7 transposon system which inserts a cassette, flanked by the Tn7L and Tn7R sequences, downstream of the *glmS* gene [121]. The method requires the double transformation, of the strain to be engineered, with a plasmid carrying the insertion cassette and a plasmid carrying the transposition machinery, TnsABCD. This was done by electroporation. The plasmids both had an R6K origin of replication was is incapable of replicating in strains that do not express the *pir* gene. This means that the plasmids are lost quickly and one can select for successful chromosomal insertion with antibiotics. Successful insertion was also checked by PCR with one primer annealing within the chromosomal glmS gene and one within the insert.

2.2.8 Characterisation of Genetic Circuits

Unless otherwise stated, all plasmids were transformed into *E. coli* MG1655 by electroporation for characterisation. Overnight cultures of the relevant strains were grown in supplemented M9 media with the appropriate antibiotics in a shaking incubator at 37°C and 200 rpm. After 16 hours of growth, the cultures were diluted 1:1000 into fresh supplemented M9 media with antibiotics and grown for 5 hours in a shaking incubator at 37°C and 200 rpm. After 5 hours the cultures were diluted 1:100 into fresh supplemented M9 media with antibiotics and induced with the relevant concentration of inducer. 200 µL of each of the induced culture were then pipetted into a clean 96 well microtitre plate and sealed with a Breathe-Easy sealing membrane. The plate was immediately placed in a plate reader and run with a program that recorded absorbance at 540 nm and fluorescence with excitation at 480 nm and emission at 540 nm every 10 minutes for 20 hours. It was incubated at 37°C and had constant double orbital shaking while not taken measurements. After completion of the plate reader program, the microtitre plate was removed from the plate reader and 1 µL of culture from each well was used to inoculate 200 µL of PBS in a clean round-bottom 96 well microtitre plate. This plate was used to take flow cytometry recordings as detailed in Section 2.1.2. Data from the flow cytometer was analysed as detailed in Section 2.1.3.

2.2.9 Characterisation of Genetic Circuits in PBS

In order to ascertain circuit function at a constant cell density some characterisation was carried out in PBS rather than media. The protocol was identical to that carried out above except that before the 1:100 dilution into M9 media, the cultures were washed twice with PBS before being diluted 1:100 into fresh PBS. These 1:100 dilutions were then serially diluted 1:2 into PBS. All of these dilutions were then induced with the relevant concentrations of inducer and pipetted into the microtitre plate as before. The plate reader recordings were only carried out for 6 hours.

2.2.10 Agar Plate Spot Inhibition Assay

Cultures of bacteriocin producing strains were grown overnight in supplemented M9 media without antibiotics but with the recorded concentration of inducer. A culture of a bacteriocin sensitive strain, *E. coli* MG1655 unless otherwise stated, was also grown overnight. Using a sterile swab, a lawn of sensitive bacteria was spread on an antibiotic free LB-agar plate. This was allowed to dry at room temperature for an hour. 3 µL of each culture was spotted on to the surface of the lawn and the plate was allowed to dry for a further hour. The plate was placed in an incubator at 37°C for 20 hours. The plate was visualised in a BioRad GelDoc using the epi-white light source.

2.2.11 Chemostat Competition Experiment

The SPoCK strain was spread on LB agar plates containing kanamycin and chloramphenicol, from glycerol stocks using an inoculation loop. The EcM-GmRed strain was spread on LB agar plates containing gentamicin, from glycerol stocks using an inoculation loop. The plates were grown overnight at 37°C. Single colonies were picked and inoculated into 5 mL of selective LB media in a 15 mL Falcon tube. The tubes were incubated overnight at 37°C and 200 rpm in a 2.5 cm orbital shaking incubator.

The Eppendorf DASbox was prepared as per the manufacturers instructions. The feed media was M9 media with glycerol as the sugar source and 0.2% arabinose. All four chambers were filled with 150 mL of media. The overnight cultures of killer and competitor were mixed together and 2 mL was used to inoculate each chamber. After 6 hours of growth, the chambers had reached an optical density of ~ 1. At this point the dilution pumps were switched on; chambers 1 and 2 were diluted at 15 mL h^{-1}, chambers 3 and 4 at 7.5 mL h^{-1}. At sample timepoints 1 mL of culture was taken from each chamber and the optical density was measured. 2 µL of the sample was used to inoculated 100 µL of PBS and the single cell fluorescence was measured in the flow cytometer.

Table 2.5: Plasmids used in this chapter.

Plasmid	Description	Source
pSEVA246	hybrid pRO1600/ColE1 ori, KnR, *luxCDABE*	SEVA collection
pSF_SC101_OXB20-daGFP	SC101 ori, KnR, OXB20 → daGFP, MCS	Oxford Genetics
pKG1022	A plasmid carrying hok/sok from pPR633.	[124]
pREG531	A plasmid carrying axe/txe from pRUM.	[125]
pHK11	BR322 ori, ApR, microcin-V	[119]
p28_OXB20	pSF_SC101_OXB20-daGFP with the ori swapped for pUC	this work
p28_OXB20-HS	p28_OXB20, hok/sok	this work
p28_OXB20-AT	p28_OXB20, axe/txe	this work
p28_OXB20-MCC	p28_OXB20, microcin-V	this work
p28_OXB20-CVA	p28_OXB20, microcin-V without transport genes	this work
p24_help-Lux	pSEVA246 with phelp promoter upstream of *luxCDABE*	this work
p24_help-Lux-HS	p24_help-Lux, hok/sok	this work
p24_help-Lux-AT	p24_help-Lux, axe/txe	this work
p24_help-Lux-MCC	p24_help-Lux, microcin-V	this work
pTNS1	Tn7 transposase expression	[121]
pUC18T-mini-Tn7T-Gm-dsRedExpress	GmR mini-Tn7 vector, mobilizable, for RFP tagging bacteria	[126]

2.3 Methods for Plasmid Stability Experiments

A number of plasmids were created for the work in Chapter 5. These plasmids are listed in Table 2.5. All plasmid sequences were confirmed with Sanger sequencing. Where the created plasmids have used SEVA vectors, the SEVA naming convention for antibiotic resistance and origin of replication has been adhered to, though the "SEVA" prefix has been removed as the plasmids don't necessarily adhere to the SEVA standards. The SEVA convention for naming of the "cargo" region has not been adhered to as the extensibility of this naming convention is limited. The cloning procedures for the construction of some of these plasmids are detailed below.

The bacterial strains used in Chapter 5 are detailed in Table 2.6. Chemically

Table 2.6: Bacterial strains used in this chapter.

Strain	Genotype	Source
E. coli NEB 5-alpha	*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 ϕ80Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	New England BioLabs
EcN	*E. coli* Nissle 1917	Prof. Ian Henderson, UK
EcN-Red	*E. coli* Nissle 1917 with chromosomal dsRedExpress and gentamicin resistance	this work
EcN-Lux	*E. coli* Nissle 1917 with chromosomal *luxCDABE* and erythromycin resistance	[13]

Table 2.7: Primers used in this chapter. Underlined sequences show restriction sites added to the amplified fragments. Upper case letters show the annealing sequence.

Primer	Sequence
P.SacI.mccV.FOR	tgacgcgagctcTGCCCTTCCCTAGAGAATCC
P.XbaI.mccV.REV	actgtctctagaGGGTCAGTGCAGAATTTTTA
P.SacI.cvaC.FOR	tgtcgcgagctcACGAGCCTTATACCGGAAA
P.HS.SacI.F	atctgagagctcTCCGGCCGAACAAACTCC
P.HS.BamHI.R	atgtcaggatccAAGGAGAAAGGGGCTACCG
P.glmS.F	CGCTGCAACTACTGGCTTA
P.Tn7.R	CACAGCATAACTGGACTGATTTC

competent cells were produced using a standard protocol. The primers used for PCR amplifications are detailed in Table 2.7. The annealing sequences were designed using Primer3Plus [123] and annealing temperatures were calculated using the NEB Tm calculator (http://tmcalculator.neb.com). The analysis of the flow cytometry followed the pipelines detailed in Section 2.1.3.

2.3.1 Chromosomal Insertion of Antibiotic Resistance and Fluorescence Reporter in to *E. coli* Nissle

The plasmid loss experiments could not be carried out using antibiotic selection for plasmid-borne resistance genes as this would prevent plasmid loss. Therefore, in order to prevent external contamination during prolonged experiments, an antibiotic resistance gene, along with a fluorescence reporter, was inserted in to *E. coli* Nissle

using the mini-Tn7 transposon system. A plasmid carrying the transposon machinery, pTNS1, and a plasmid carrying the sequence to be inserted, pUC18T-mini-Tn7T-Gm-dsRedExpress, were transformed in to electrocompetent *E. coli* Nissle by electroporation. Correct insertion was then tested by colony PCR using primers P.Tn7.R and P.glmS.F. Unfortunately, absence of the two plasmids was not checked for until after 25 days of a plasmid-loss experiment at which point it was discovered that the pUC18T-mini-Tn7T-Gm-dsRedExpress plasmid was still present. Further attempts to cure the strain of the plasmid failed, therefore, the plasmid-loss experiment was redone in EcN-Lux.

2.3.2 Cloning of p28_OXB20 Based Plasmids

All of the below DNA manipulations were performed in *E. coli* NEB 5-alpha and followed the instruction detailed by the reagent manufacturers unless otherwise stated.

Plasmid p28_OXB20-HS was created by digesting pKG1022 with HindIII and SacI and gel extracting the 633 bp fragment. This was ligated in to p28_OXB20 which had been digested using the same restriction enzymes. Plasmid p28_OXB20-AT was created by digesting pREG531 with BamHI and SacI and gel extracting the 1390 bp fragment. This was ligated in to p28_OXB20 which had been digested using the same restriction enzymes. Plasmid p28_OXB20-MCC was created by PCR amplification of pHK11 with primers P.SacI.mccV.FOR and P.XbaI.mccV.REV to amplify the microcin-V fragment and anneal restriction enzyme recognition sites to the ends. This fragment was then digested with XbaI and SacI and a fragment at 4.8 Kbp was extracted. This was ligated in to p28_OXB20 which had been digested using the same restriction enzymes. Plasmid p28_OXB20-CVA was created by PCR amplification of pHK11 with primers P.SacI.cvaC.FOR and P.XbaI.mccV.REV to amplify the cvi-cvaC fragment and anneal restriction enzyme recognition sites to the ends. This fragment was then digested with XbaI and SacI and a fragment at 1.2 Kbp was extracted. This was ligated in to p28_OXB20 which had been digested using the same restriction enzymes. The plasmid p28_OXB20 and the inserted fragments are shown in Figure 2.9. These plasmids were transformed in to EcN-Lux

using a standard heat-shock transformation method.

2.3.3 Cloning of p24_help-Lux Based Plasmids

The p24_help-Lux based plasmids were produced in two steps; first the PSK systems were cloned in to the MCS of plasmid pSEVA246, then the phelp promoter was cloned upstream of the *luxCDABE* cassette. p28_OXB20-HS was PCR amplified with P.HS.SacI.F and P.HS.BamHI.R to amplify the hok/sok fragment and anneal restriction enzyme recognition sites to the ends. This fragment was then digested with SacI and BamHI and ligated in to pSEVA246. p28_OXB20-AT was digested with SacI and BamHI and the 1.3 Kbp fragment was gel extracted and ligated in to pSEVA246. p28_OXB20-MCC was digested with SacI, XbaI and ApaLI; the third enzyme was added to cut the unwanted fragment and enable the gel extraction of the microcin-V 4.8Kbp fragment. This was ligated in to pSEVA246 which had been digested with SacI and XbaI. The plasmid pSEVA246 and the inserted fragments are shown in Figure 2.10. The phelp promoter was then cloned upstream of the *luxCDABE* cassette for the PSK bearing plasmids and the pSEVA246 using Gibson Assembly [127] by Anjali Doshi (Danino Lab, Columbia University, USA) These plasmids were transformed in to EcN using a standard heat-shock transformation method.

2.3.4 Growth Rate Assays - p24_help-Lux Based Plasmids

Two changes had to be made to the general protocol when recording growth curves for the luminescent plasmids. Firstly, they had to be grown in selective media to prevent the rapid plasmid loss in p24_help-Lux. Secondly, absorbance measurements were additionally taken at 410 nm due to the frequency of the luminescence.

2.3.5 Plasmid Stability Assays in Liquid Culture

Assays were performed in LB media in sterile deep square 96-well (2.2 mL) polypropylene plates. The edge wells were used for LB media controls so that the strains weren't affected by any potential plate edge conditions [128]. The 96-well plate was split into two sections; one section with 500 mL LB media containing erythromycin (100 µg/mL) and kanamycin (25 µg/mL) and the second with LB media

containing only erythromycin The erythromycin was used to prevent external contamination as the resistance gene is encoded on the chromosome of the EcN-Lux bacterial strain. The kanamycin is used to ensure plasmid maintenance, generating a plasmid-bearing control throughout the experiment.

Three double selective wells for each strain were inoculated by picking single colonies from selective LB agar plates. Six single selective wells were inoculated with EcN-Lux, as a plasmid-free control, by picking single colonies from selective LB agar plates. The plate was covered with a System Duetz breathable silicon sandwich (EnzyScreen, The Netherlands) and clamped down in a shaking incubator at 37°C and 350 rpm for 24 hours. This produced the time point 0 cultures. The procedure is shown in Figure 2.11.

These cultures were then used to inoculate 1:500 a fresh plate, arranged with the same selective layout. Each double selective cultures were used to inoculate 3 replicates in single selective media. This produced 9 replicates of each plasmid-bearing strain in single selective media and 3 in double selective media. The EcN-Lux plasmid-free control had 6 replicates in single selective media. Although the double selective media was inoculated with the EcN-Lux strain, as expected, no bacteria grew. This plate was covered and incubated as before. For subsequent passages the plate was replicated 1:500 in to fresh media with a one-to-one mapping between wells.

At each passage, 200 µL of the overnight culture was transferred in to a black clear-bottom 96-well microtitre plate (Greiner Bio-one, Germany) and absorbance (540nm), fluorescence (ex=480 nm, em=540 nm), and luminescence were measured in a microplate reader. In addition, 1 µL of the overnight culture was transferred in to a round-bottom 96-well microtitre plate containing 200 µL of PBS and the single cell fluorescence was measured using a flow cytometer.

The assay for the luminescent plasmids differed slightly from that described above for the fluorescent plasmids. The assays could not be performed with erythromycin as the EcN bacterial host did not contain the resistance gene. In order to prevent external contamination in LB cultures without any antibiotic, the entire

plate was additionally exposed to UV light for 40 minutes before inoculation at each passage [129]. Further, only absorbance and luminescence could be measured so the preparation of the plate for flow cytometry was not carried out. Absorbance was additionally measured at 410 nm.

2.3.6 Luminescent Plasmid Stability Assays *in vivo*

This work was carried out by Anjali Doshi and others in the research group of Tal Danino (Columbia University, USA).

All animal work was approved by the institutional committee on animal care (Columbia, AC-AAAN8002). The protocol requires animals to be euthanised when tumours reach 2 cm^3, or under veterinary staff recommendation. The cell line (MC26-LucF, Tanabe laboratory, Massachusetts General Hospital) was obtained from, and authenticated by, the Tanabe laboratory, MGH. The cell line was tested to be mycoplasma-free before implantation in mice. Sample sizes for mice were determined by expected effect size to produce a power of 0.8-0.9. Mice were blindly randomized into various groups using a random number generator.

Animal experiments were performed on 6-week-old female BALB/c mice (Taconic Biosciences) with bilateral subcutaneous hind flank tumours from an implanted mouse colon cancer cell line (MC26-LucF). The concentration for implantation of the tumour cells was 10^8 cells per ml in DMEM (no phenol red). Cells were then implanted subcutaneously at a volume of 100 µL per flank, with each implant consisting of 10^7 cells. The MC26 cells where given 14 days to graft and grow until they reached a size of approximately 150-200 mm^3. Along with an EcN-Lux strain as a control, EcN:p24_help-Lux, EcN:p24_help-Lux-HS, EcN:p24_help-Lux-AT and EcN:p24_help-Lux-MCC were grown up in LB with kanamycin until exponential phase, washed three times in sterile PBS and then injected intravenously at a dosage of $1x10^6$ bacteria in 100 µL. After bacterial colonisation of the tumours, animals were monitored via IVIS (IVIS 200, Caliper Life Sciences) to detect the presence of the plasmid via bioluminescence. Briefly, this involved anaesthetising the mice with isoflurane and imaging with an open filter on the auto-exposure setting. After 7 days, each of the tumours were sterilely extracted and homogenised

using a tissue dissociator (Miltenyi), an aliquot of which was seeded on each of LB plates and LB with kanamycin plates. The ratio was calculated by comparing the mean counts of plasmid bearing colonies on LB plates and LB plates with kanamycin from each tumour [13].

2.3.7 Strain Dilution Assays

Overnight cultures of each strain were grown, from single colonies picked from selective LB agar plates, in 5 mL of selective LB media. Different dilutions of plasmid-bearing to plasmid-free strains were then produced by pipetting different volumes of plasmid-bearing overnight culture in to EcN-Lux overnight culture in percentages from 10% through to 100%, keeping a total volume of 100 µL. For example, to produce a 30% culture 30 µL of plasmid-bearing overnight culture was added to 70 µL of EcN-Lux overnight culture. Note that the overnight cultures were not normalised to the same optical density, so a 30% culture does not mean that 30% of the cells are plasmid-bearing. Assays were performed in LB media in sterile autoclaved deep square 96-well (2.2 mL) polypropylene plate. Each well was filled with 500 µL of LB media with erythromycin (100 µg/mL). The wells were then inoculated in a random pattern with 3 replicates of each of the diluted cultures. The deep well plate was then sealed and incubated as above. At the same time a 96-well microtitre plate was prepared with 200 µL of PBS in each well and inoculated from the same diluted cultures and measured using the same flow cytometry set up as above. After 24 hours the plate was taken from the incubator and passaged, as above, while a flow cytometry plate was also prepared and measured. This was carried out for a total of 2 passages. The data from passage 1-2 was used for the model fitting as this allowed the cells time to get used to growing in the deep-well plate. However, for the MCC strain it was necessary to use data for passage 0-1 because after 24 hours all plasmid-free bacteria had been killed.

2.3.8 Plasmid-bearing Cell Clustering

After the flow cytometry data was processed to select only the singlet bacteria, cells were classified as plasmid-bearing or plasmid-free. This classification was achieved

by clustering on the GFP fluorescence measurement as only plasmid-bearing cells are still producing GFP. Two methods for clustering were tested; one dimensional mixture models and a minimum density estimator.

The mixture model method is equivalent to that used to remove debris in the initial flow cytometry processing. Mixture models of one and two clusters are fitted to the data and integrated completed likelihood (ICL) [115] is used to determine which model fits the data best. If two clusters are found, the one with the higher fluorescence level is classified as the plasmid-bearing population and the other as plasmid-free. If only one cluster is found, the cluster is classed as plasmid-bearing if the median is above a user defined threshold. The minimum density estimator fits a kernel density estimate to the data and identifies the two largest peaks. A threshold is drawn at the minima between those two peaks and the cluster at the higher fluorescence level is classed as plasmid-bearing.

Neither of these methods works particularly well when one of the peaks is very small. As such, it was determined that a user defined threshold produced more reliable results, particularly at the key experimental points when the plasmid-free population is just establishing.

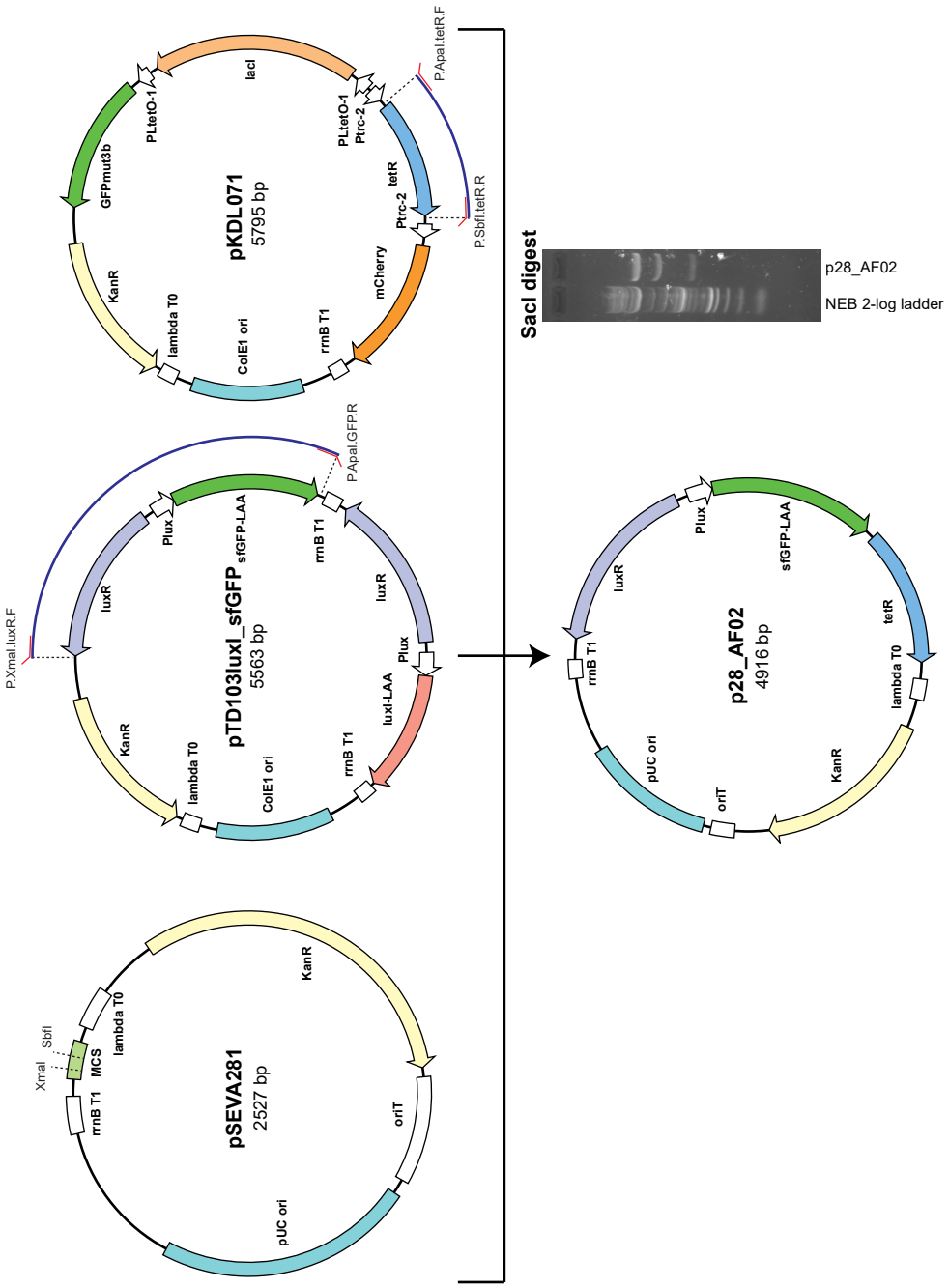

Figure 2.3: Cloning to produce p28_AF02.

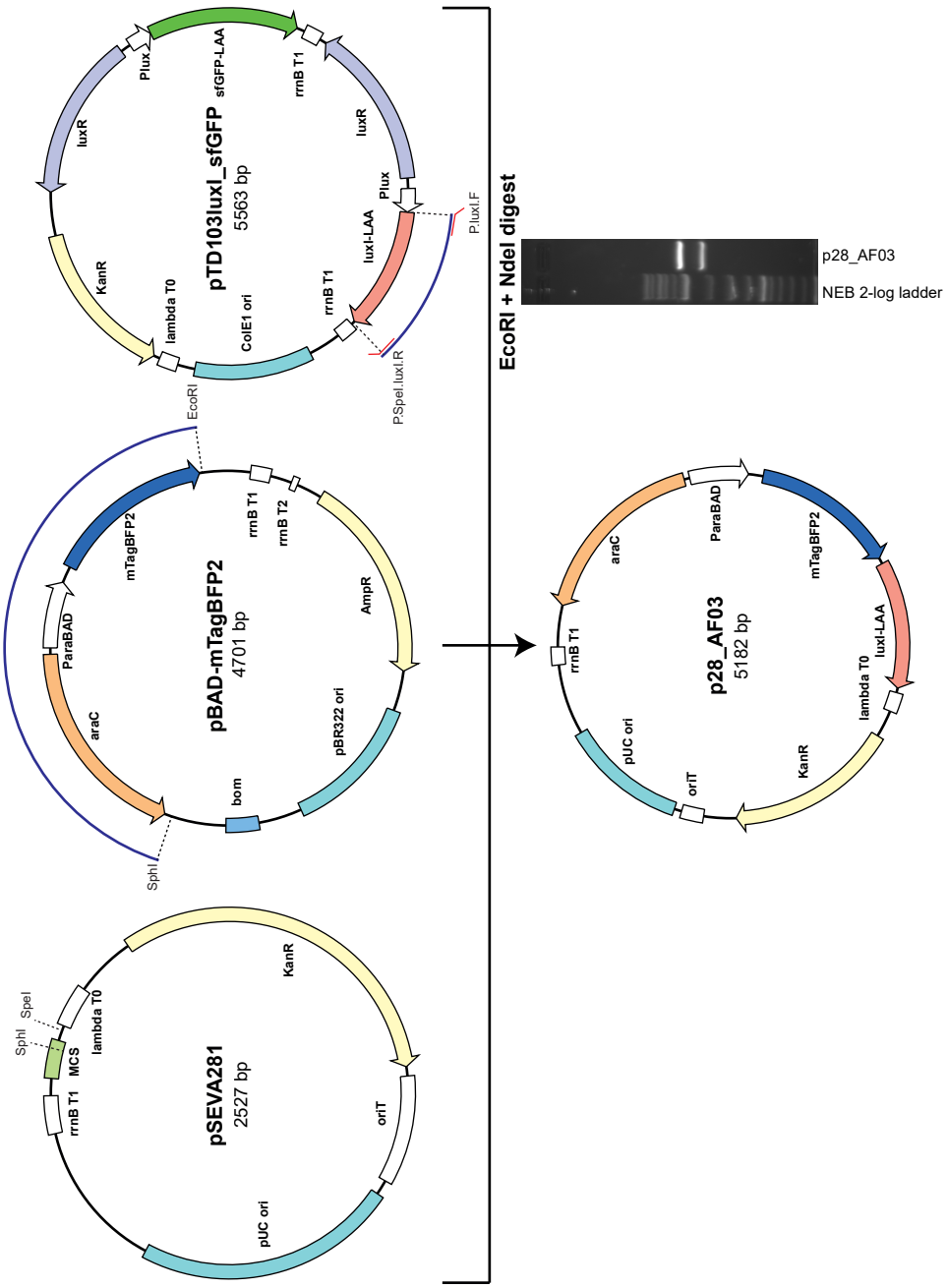

Figure 2.4: Cloning to produce p28_AF03.

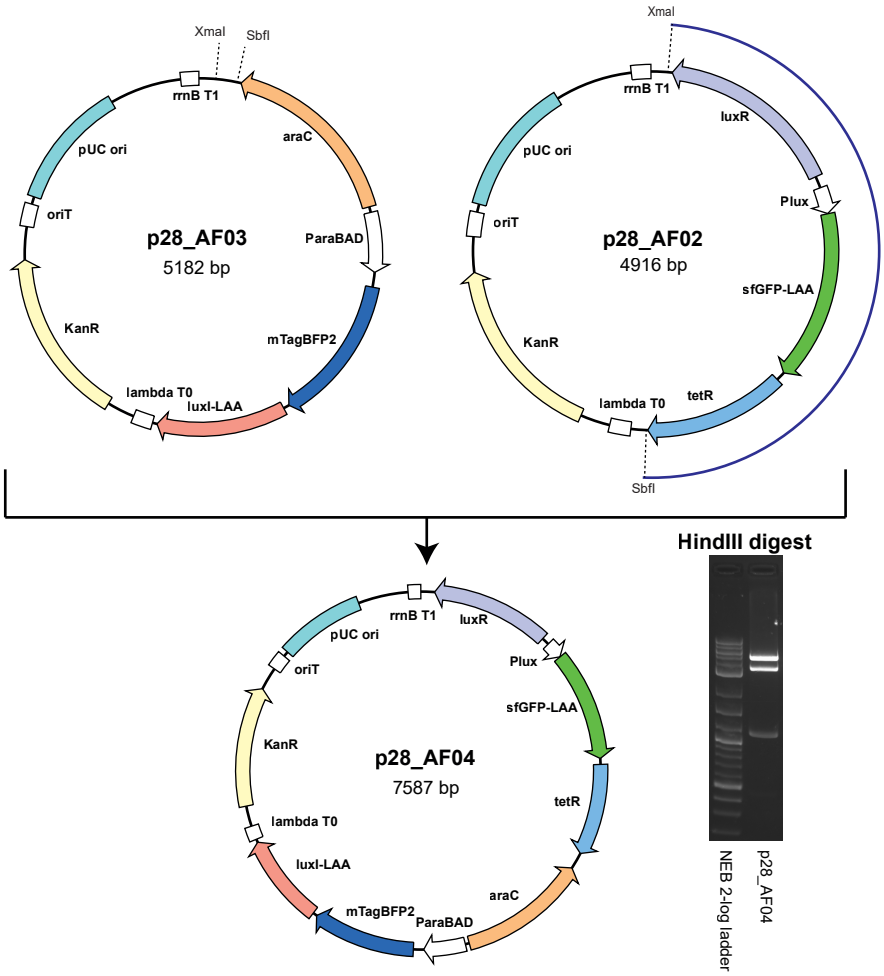

Figure 2.5: Cloning to produce p28_AF04.

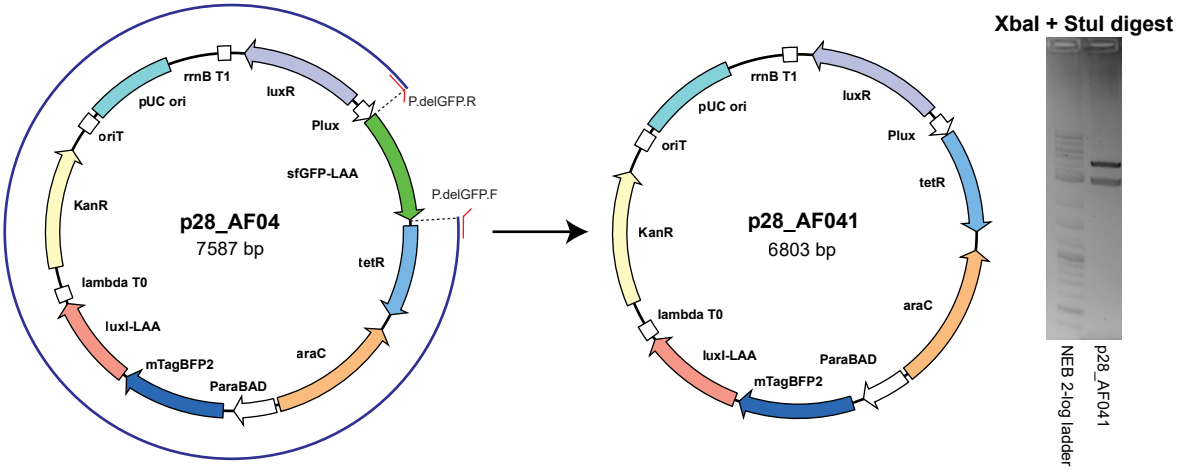

Figure 2.6: Cloning to produce p28_AF041.

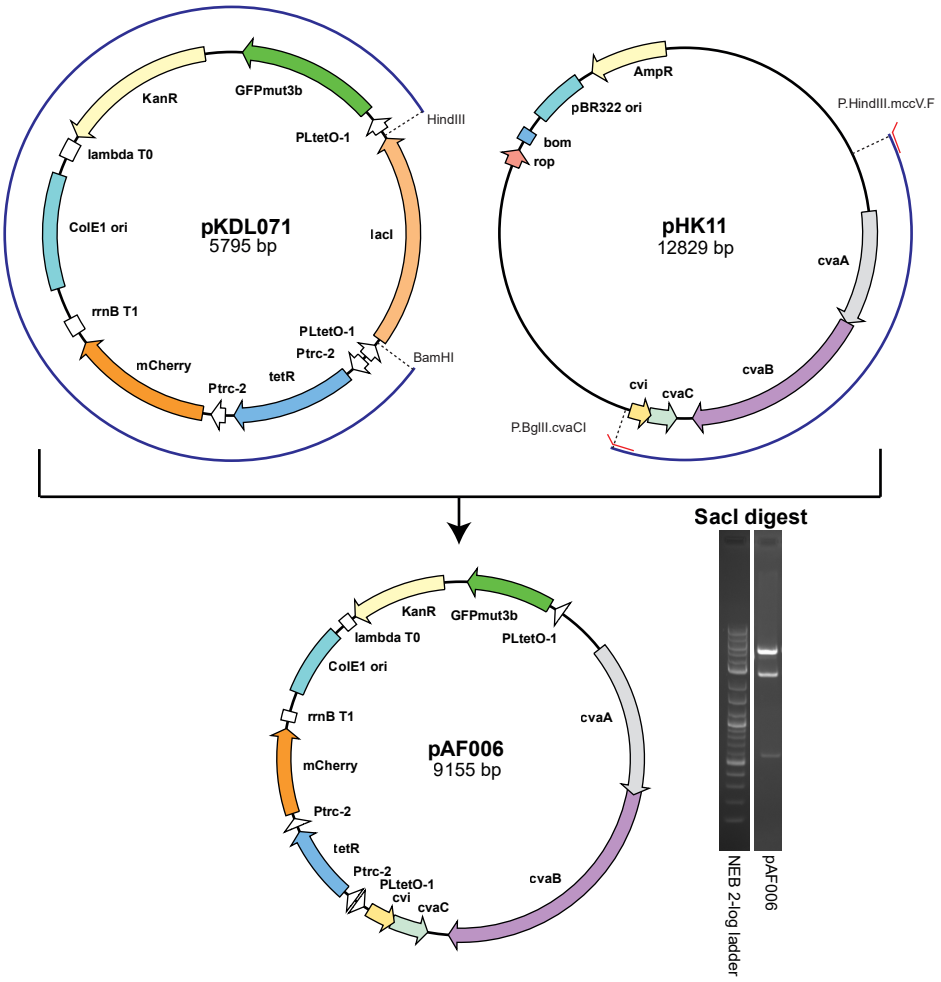

Figure 2.7: Cloning to produce pAF006.

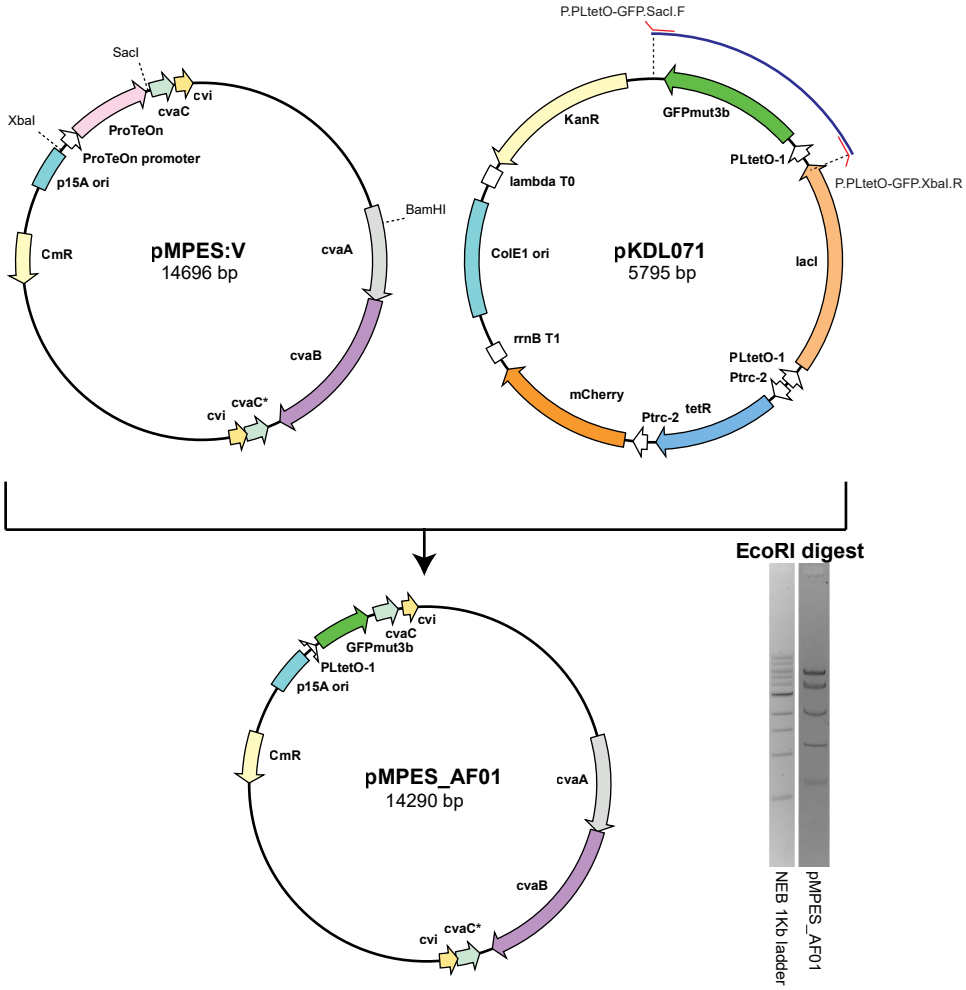

Figure 2.8: Cloning to produce pMPES_AF01.

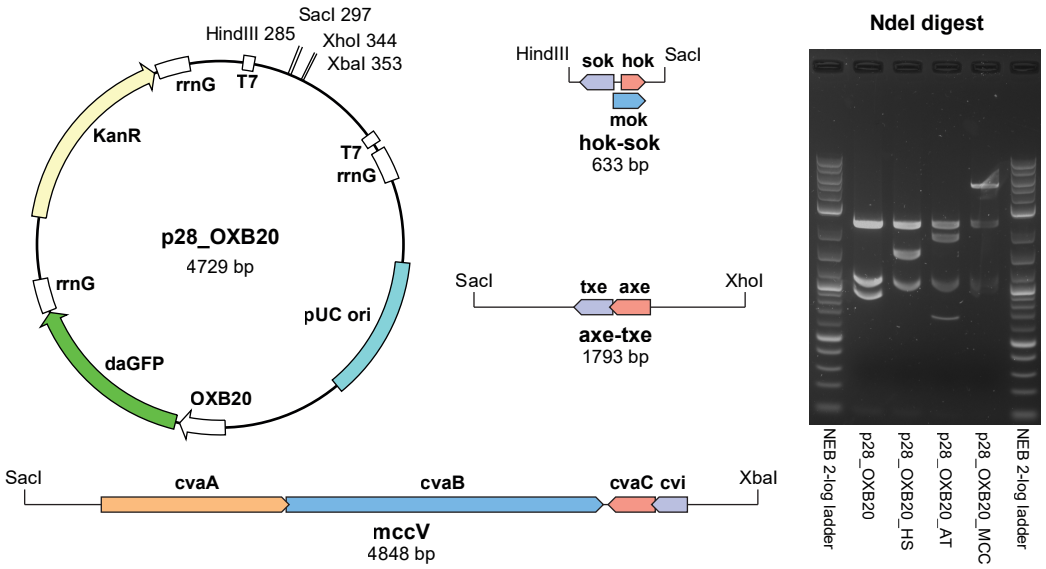

Figure 2.9: Cloning of p28_OXB20 based plasmids. Restriction enzyme cut sites for the insertion of the TA and bacteriocin fragments are marked on the plasmid.

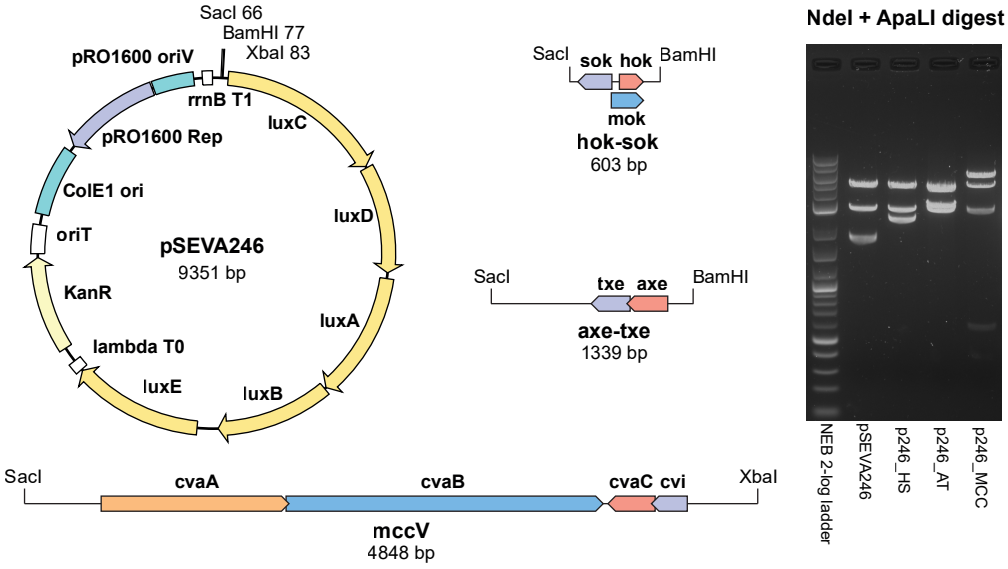

Figure 2.10: Cloning of p24_help-Lux based plasmids. Restriction enzyme cut sites for the insertion of the TA and bacteriocin fragments are marked on the plasmid.

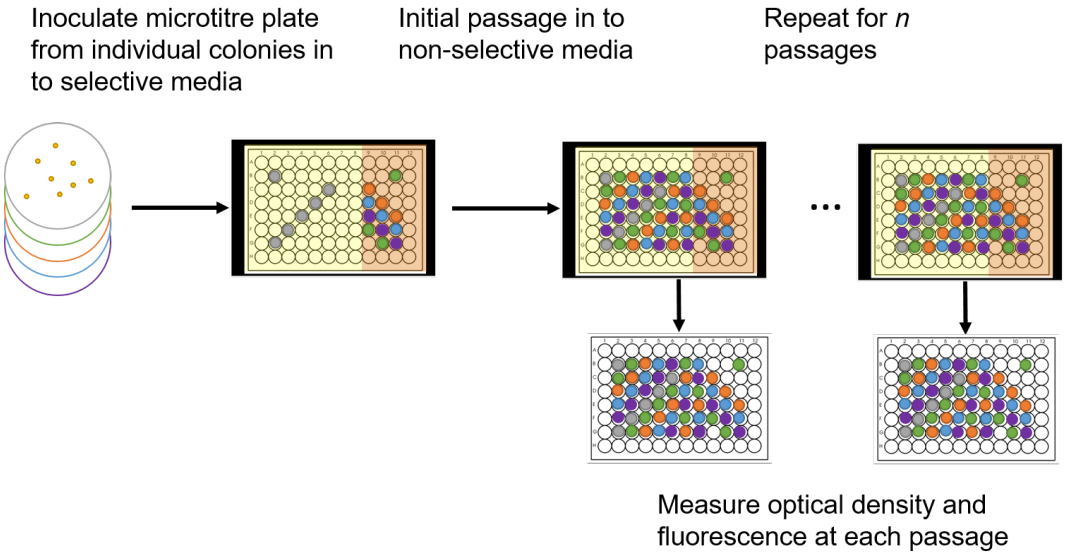

Figure 2.11: Inoculation layout and passage procedure. Strains are streaked from glycerol stocks on to LB agar plates with appropriate antibiotics. Three colonies of each strain are used to inoculate a 96-well plate; the plasmid-bearing strains are inoculated in to wells containing kanamycin and erythromycin (orange - LB+Kn+Er), the EcN-Lux controls are inoculated in to wells containing only erythromycin (yellow - LB+Er). The plate is grown for 24 hours after which a fresh 96-well plate is inoculated from the overnight cultures; the plasmid-bearing strains are used to inoculate three replicate (LB + Er) wells from each (LB + Kn + Er) biological replicate. All wells from the overnight plate are sampled, with optical density, fluorescence and luminescence being recorded. The newly inoculated plate is grown for another 24 hours after which it is passaged and sampled.

Chapter 3

A System for Stabilising Competing Populations of Bacteria

Competition between heterogeneous populations of engineered bacteria or engineered and wild type strains is a topic that is becoming increasingly relevant with the expansion in the complexity of synthetic circuits and the environments in which they are used. Increasing system complexity implies increasing burden, making engineered strains less competitive. In addition, intermediary products in pathways can be toxic to certain strains of microbes, further reducing fitness. Division of labour between bacteria will become necessary. However, as soon as there are multiple sub-populations, with different growth rates, within an environment, competition between those populations becomes a challenge. Thus, one has to overcome the competitive exclusion principle, which states that two species competing for the same resource cannot coexist [130]. This is because, even with a very slight difference in fitness, the stronger competitor will dominate leading to the extinction, or adaptation to a new niche, of the weaker species. The principle should, perhaps, include the caveat: "in stable environments and in the absence of other interactions". This caveat may help to explain supposed problems with the competitive exclusion principle such as the "paradox of the plankton" [131, 132].

Wild bacteria live in complex communities [133] with mutualistic and competitive interactions producing complex dynamics [134]. Previous attempts to design synthetic microbial communities have relied on spatial segregation [96] or mutual-

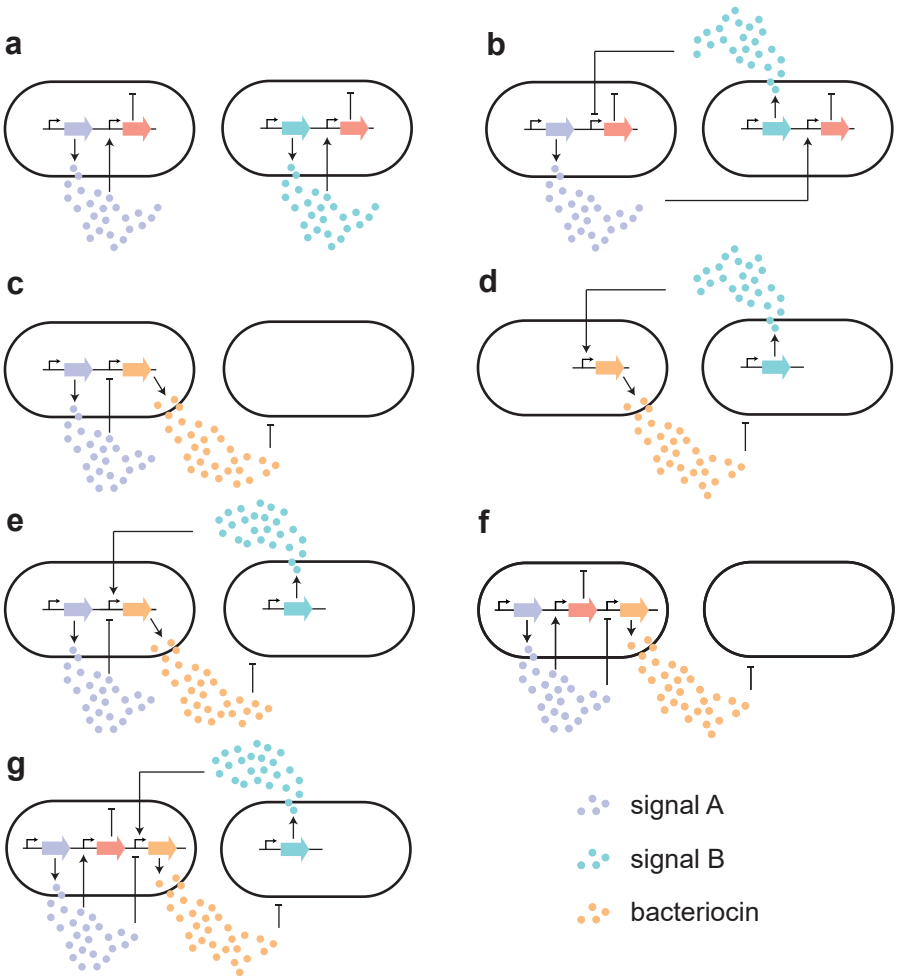

Figure 3.1: Designs for synthetic community control systems.

ism [95] to maintain multiple sub-populations. The control of the density of single species populations has been achieved through the use of quorum sensing to control self-killing [135] and more recently this has been extended to a two species system [102], as in Figure 3.1a. More complex predator-prey systems have also been developed which produce oscillatory populations of two strains [100] using a design similar to that shown in Figure 3.1b. These systems involve the engineering of all strains within the community. However, this requirement may not be desirable in industrial settings and is clearly not possible when working in natural environments such as the human gastrointestinal tract.

With the restriction that only one of the strains within a synthetic community can be engineered, new approaches need to be taken. Bacteriocins, as described

previously, are a competition mechanism that is widely used by wild-type bacteria in complex environments such as the human gastrointestinal tract [136, 137, 138]. These bacteriocins can be harnessed to reduce the fitness of a competitor strain in a community, while allowing the producer to continue growing. Indeed they have been used in engineered systems to eliminate pathogenic strains [88]. If the assumption is made that the engineered strain will grow more slowly than the competitor strain, bacteriocins can be produced while the engineered strain has a low population density, producing a niche for it to grow into. Once the population has grown above a desired threshold, the bacteriocin can be turned off and the competitor can grow again. A cartoon of this system is shown in Figure 3.1c. This is the inverse of a system discovered in *Burkholderia thailandensis* in which the production of a bacteriocin was induced by an endogenous quorum sensing system [139, 140]. A slight modification to this system would be to sense a native signal from the competitor, though this would restrict the potential environments into which the system could be introduced. This would allow the engineered strain to turn on bacteriocin production only when the competitor population is high, Figure 3.1d. Combining both of these systems would limit bacteriocin production to states in which the engineered strain's density is low and the competitor's density is high, preventing the extermination of the competitor population when both densities are low, Figure 3.1e.

With a further modification to the bacteriocin producing system, the assumption that the competitor grows faster can be removed. The addition of a quorum controlled self-killing mechanism to the engineered strain would place an upper limit on the engineered strain's density, Figure 3.1f. This, in addition to the lower limit produced by the bacteriocin system, would ensure niches are available for both strains while only requiring the engineering of a single strain. The combination of all of these constituent parts, Figure 3.1g, would ensure the viability of the competitor population regardless of its relative fitness.

In this Chapter I will show, using mathematical models, that the system shown in Figure 3.1c, which from here will be called Stabilised Populations by Commu-

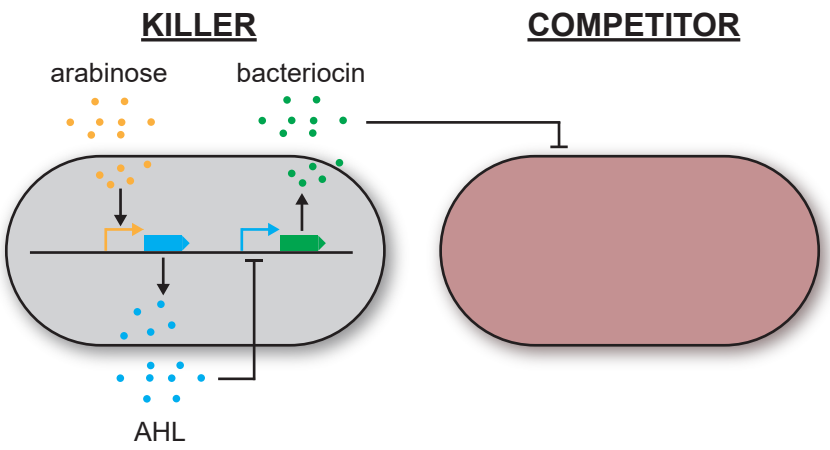

Figure 3.2: A simple cartoon of the SPoCK system.

nity Killing (SPoCK), is capable of producing both stable and oscillating populations dependent on parameters that can be controlled in a chemostat environment. In the following Chapter I will construct an example of this system and through characterisation experiments and Bayesian computational methods, determine the behaviours that this instantiation is capable of.

3.1 A Model of SPoCK in Chemostat

The SPoCK system can be described as a dynamic competition mechanism in which the mechanism for reducing the fitness of competitor is controlled dynamically dependent on the population density of the engineered strain. A simple schematic of an implementation of the SPoCK system is shown in Figure 3.2. The bacteriocin secreted by the killer strain kills the competitor strain, allowing the killer to outcompete it. The bacteriocin expression is turned off by a quorum sensing molecule produced by the killer strain. This results in the bacteriocin being expressed only when the killer population is low, and turned off when the population is high. In this implementation, the production of the quorum sensing molecule is placed under the control of an inducible promoter. This allows for the tuning of the rate of production of the quorum sensing molecule, which allows for exogenous control of the population density at which the bacteriocin production will be switched off.

The system shown in Figure 3.2 can be described using a mathematical model. In this simple model, the two bacterial populations grow at different rates which are dependent on the concentration of a substrate for which they compete. The environment has a constant dilution rate which removes cells and molecules from the environment but introduces fresh substrate. Finally, the competitor strain is susceptible to the bacteriocin produced by the killer strain which results in a death rate of competitor dependent on bacteriocin concentration. This leads to an ODE model in which the strain concentrations vary according to:

$$\frac{dX}{dt} = (\mu_X(t) - D)X(t) \tag{3.1}$$

$$\frac{dC}{dt} = (\mu_C(t) - D - \omega\,\beta(t))C(t) \tag{3.2}$$

where X is the killer strain containing the SPoCK system and C is the competitor strain. The μ are the respective growth rates, D is the dilution rate, and ω is the death rate of the competitor strain per M of bacteriocin, β. The growth rates, μ_X and μ_C, are given by Monod growth laws:

$$\mu_X(t) = \frac{\mu_{X_{max}} S(t)}{K_X + S(t)} \tag{3.3}$$

$$\mu_C(t) = \frac{\mu_{C_{max}} S(t)}{K_C + S(t)} \tag{3.4}$$

where the μ_{max} are the respective maximal growth rates, S is the substrate concentration, and K_X and K_C are the half saturation constants. The concentration of substrate, S, varies according to:

$$\frac{dS}{dt} = D(S_0 - S(t)) - \frac{\mu_X(t)X(t)}{\gamma_X} - \frac{\mu_C(t)C(t)}{\gamma_C} \tag{3.5}$$

where S_0 is concentration of substrate in the feed media, and the γ are the respective yield coefficients.

The concentration of bacteriocin varies according to:

$$\frac{d\beta}{dt} = k_\beta(t)X(t) - D\beta(t) \tag{3.6}$$

where k_β is the rate of bacteriocin production. This rate is inversely dependent on the concentration of AHL in the environment. This relationship is described using the inverse Hill function:

$$k_\beta(t) = \frac{k_{\beta_{\max}} K_\beta^{n_\beta}}{K_\beta^{n_\beta} + A(t)^{n_\beta}} \tag{3.7}$$

where $k_{\beta_{\max}}$ is the maximal expression rate of bacteriocin, K_β is the concentration of AHL at which bacteriocin is half-maximally expressed and n_β is the Hill coefficient. In this simple model the expression rate of AHL is modelled as constitutive at a rate k_A that would conceptually be varied by changing the arabinose concentration in the media. A further simplification in this model is that k_A is not considered to be time dependent as it is assumed that the concentration of arabinose in the media would not vary. However, under real world conditions, if another preferable sugar source was not present and the arabinose was not being replenished quickly through dilution, the arabinose may be consumed and the concentration would decrease. The concentration of AHL, therefore, varies with time according to:

$$\frac{dA}{dt} = k_A X(t) - DA(t) \tag{3.8}$$

Some values of the parameters can be found in the literature or given sensible values based on similar systems, Table 3.1.

3.1.1 The SPoCK System at Steady-State

By analysing the system at steady state it is possible to determine whether there are any conditions under which the two strains can reach an equilibrium and co-exist

Table 3.1: Model parameters for the simple chemostat model.

Parameter	Description	approx. value	unit
D^*	Dilution rate	0-$\mu_{C_{max}}$	h^{-1}
${S_0}^*$	Reservoir substrate concentration	$1-4$	gL^{-1}
${k_A}^*$	Expression rate of AHL from killer strain	$\sim 10^{-21}$	mol $cell^{-1}$ h^{-1}
$\mu_{X_{max}}$	Maximum growth rate of killer strain	0.4-3	h^{-1}
$\mu_{C_{max}}$	Maximum growth rate of competitor strain	0.4-3	h^{-1}
γ_X, γ_C	Yield coefficient of killer and competitor strains	10^{12}	cells g^{-1}
K_X, K_C	Concentration of substrate at which killer and competitor strains growth is half-maximal	2	gL^{-1}
ω	Death rate of competitor by bacteriocin	$\sim 10^8$	M^{-1} h^{-1}
$k_{\beta_{\max}}$	Maximum expression rate of bacteriocin from killer strain	$\sim 10^{-22}$	mol $cell^{-1}$ h^{-1}
K_β	Concentration of AHL at which bacteriocin expression is half-maximal	$\sim 10^{-7}$	M
n_β	Hill coefficient of bacteriocin production	2	-

* Parameters are easily experimentally controllable.

stably.

When the killer strain X is at steady state:

$$\begin{aligned} \dot{X} &= (\mu_X - D)X = 0 \\ \mu_X &= D \end{aligned} \tag{3.9}$$

Using Equation 3.3:

$$\begin{aligned} D &= \mu_X \\ &= \frac{\mu_{X_{max}} S}{K_X + S} \\ S &= \frac{D\, K_X}{\mu_{X_{max}} - D} \end{aligned} \tag{3.10}$$

In the above Equation S is constant, since the right-hand-side (RHS) are all constants. As such, the nullcline for $dX/dt = 0$ is a point in S-space. Further, since S must be constant:

$$\begin{aligned} \dot{S} &= D(S_0 - S) - \frac{\mu_X\, X}{\gamma_X} - \frac{\mu_C\, C}{\gamma_C} = 0 \\ S &= S_0 - \frac{\mu_X\, X}{D\ \gamma_X} - \frac{\mu_C\, C}{D\ \gamma_C} \end{aligned} \tag{3.11}$$

The μ_X and μ_C can be replaced with the RHS's of Equations 3.3 and 3.4 and the S's can be replaced with the RHS of Equation 3.10. The subsequent Equation can be rearranged to give solutions for X or C as a function of the other:

$$X = \gamma_X\left(\frac{D\ K_X}{D - \mu_{X_{max}}} + \frac{C\ K_X\ \mu_{C_{max}}}{D\gamma_C K_C - D\gamma_C K_X - \gamma_C K_C \mu_{X_{max}}} + S_0\right) \tag{3.12}$$

$$C = \frac{\gamma_C(D(K_C - K_X) - K_C\ \mu_{X_{max}})(\frac{D\ K_X}{\mu_{X_{max}} - D} - S_0 + \frac{X}{\gamma_X})}{K_X\ \mu_{C_{max}}} \tag{3.13}$$

When the competitor strain C is at steady state:

$$\begin{aligned} \dot{C} &= (\mu_C - D - \omega\beta)C = 0 \\ \mu_C &= D + \omega\beta \end{aligned} \tag{3.14}$$

Using Equation 3.4:

$$\begin{aligned} D + \omega\beta &= \mu_C \\ &= \frac{\mu_{C_{max}} S}{K_C + S} \\ \beta &= \frac{\mu_{C_{max}}\ S}{\omega(K_C + S)} - \frac{D}{\omega} \end{aligned} \tag{3.15}$$

The nullcline for C at steady state is therefore represented by a line on the β vs S plane, shown in Figure 3.3.

Equation 3.15 gives the relationship between the bacteriocin concentration β and the substrate concentration S when C is at steady state and Figure 3.3 shows

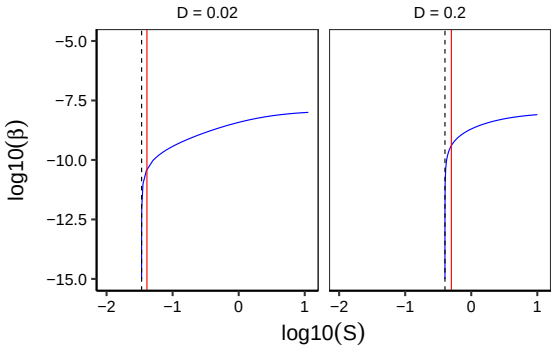

Figure 3.3: Nullclines of $dX/dt = 0$ (Equation 3.10, red line) and $dC/dt = 0$ (Equation 3.15, blue line) at two different dilution rates on the S versus β plane. The dashed line shows the concentration of substrate below which the competitor strain cannot grow faster than the dilution rate. ($K_C = K_X = 2,\ \mu_{X_{max}} = 1,\ \mu_{C_{max}} = 1.2,\ \omega = 10^8$)

the intersection of the nullcline of $dX/dt = 0$ and $dC/dt = 0$ in β-S state space. The two Equations 3.10 and 3.15 can be combined to give an expression for the bacteriocin concentration β when both X and C are at steady state:

$$\beta = \frac{D(D(K_X - K_C) - K_X\ \mu_{C_{max}} + K_C\ \mu_{X_{max}})}{(D(K_C - K_X) - K_C\ \mu_{X_{max}})\omega} \tag{3.16}$$

As all of the terms on the RHS are constants, β must be a constant when X and C are at steady state, as can be seen from the single point of intercept of the nullclines in Figure 3.3. As such,

$$\begin{aligned} \dot{\beta} &= (k_\beta X - D\beta) = 0 \\ \beta &= \frac{k_\beta X}{D} \end{aligned} \tag{3.17}$$

Equations 3.16 and 3.17 can be combined, and noting the expression for k_β given in Equation 3.7, can be rearranged to give an Equation for X:

$$X = \frac{D^2 K_\beta^{-n_\beta}(D(K_X - K_C) - K_X\mu_{C_{max}} + K_C\mu_{X_{max}})(K_\beta^{n_\beta} + A^{n_\beta})}{k_{B_{max}}(D(K_C - K_X) - K_C\mu_{X_{max}})\omega} \tag{3.18}$$

This expression is in terms of X, A and a number of constants. Since X is at steady

state, and therefore constant, so must A. This gives:

$$\begin{aligned} \dot{A} &= (k_A X - DA) = 0 \\ A &= \frac{k_A X}{D} \end{aligned} \tag{3.19}$$

which can be used to replace the A in Equation 3.18. Finally, using the expression for X at steady state given in Equation 3.12, expressions for C and X when both species are at steady state can be produced. These equations are rather unwieldy so they will not be written out here but can be found in Appendix A.

The roots of the system at steady state were calculated using rootSolve in R which uses a Newton-Raphson method [141]. The Jacobian of the system at steady state, shown below, was determined using Mathematica and the eigenvalues at each root were calculated in R. The steady states were classified based on the number of negative real parts of the eigenvalues and whether they were complex or not, as with two-dimensional systems. If all five of the eigenvalues had negative real parts the steady state was classified as stable. If it had a mix of negative and positive real parts it was classified as unstable. If none of the eigenvalues were complex, the steady state was classified as a node and if there were complex eigenvalues it was classified as a foci. For all parameter values tested below, there were at least two complex eigenvalues. As such, all steady states plotted below are foci.

$$\begin{pmatrix} \frac{\mu_{X_{max}} S}{K_X+S} - D & 0 & 0 & 0 & \left(\frac{\mu_{X_{max}}}{K_X+S} - \frac{\mu_{X_{max}} S}{(K_X+S)^2}\right) X \\ 0 & \frac{\mu_{C_{max}} S}{K_C+S} - D - \beta\omega & -C\omega & 0 & C\left(\frac{\mu_{C_{max}}}{K_C+S} - \frac{\mu_{C_{max}} S}{(K_C+S)^2}\right) \\ \frac{K_\beta^{n_\beta} k_{B_{max}}}{A^{n_\beta}+K_\beta^{n_\beta}} & 0 & -D & -\frac{A^{n_\beta-1} K_\beta^{n_\beta} k_{B_{max}} n_\beta X}{\left(A^{n_\beta}+K_\beta^{n_\beta}\right)^2} & 0 \\ k_A & 0 & 0 & -D & 0 \\ -\frac{\mu_{X_{max}} S}{\gamma_X (K_X+S)} & -\frac{\mu_{C_{max}} S}{\gamma_C (K_C+S)} & 0 & 0 & \frac{\mu_{X_{max}} SX}{\gamma_X (K_X+S)^2} - \frac{C\mu_{C_{max}}}{\gamma_C (K_C+S)} + \frac{C\mu_{C_{max}} S}{\gamma_C (K_C+S)^2} - \frac{\mu_{X_{max}} X}{\gamma_X (K_X+S)} - D \end{pmatrix} \tag{3.20}$$

There are three model parameters that are easily changed in a real world experiment: the dilution rate D, the richness of the media S_0, and the expression rate of AHL from the killer strain k_A. Relative growth rates may also be changed by increasing or decreasing the burden on the engineered strain but this is not as easily

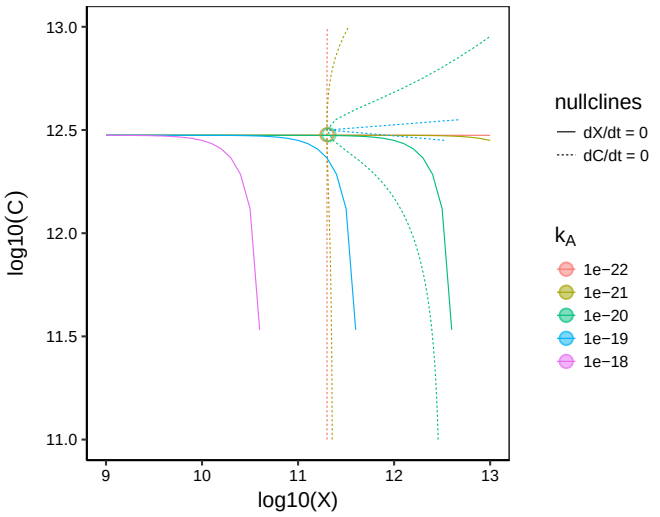

Figure 3.4: Steady state analysis of simple SPoCK in chemostat model for $S_0 = 4$ gL^{-1} and $D = 0.1$ h^{-1}. The different colours show the behaviour with different rates of AHL expression; a parameter controllable in the real system by changing arabinose concentration. The steady state concentrations of each species are shown as circles with empty circles showing unstable foci and filled circles (none present in this parameter space) showing stable foci. ($\mu_{X_{max}} = 1$ h^{-1}, $\mu_{C_{max}} = 1.2$ h^{-1})

controlled as the other parameters. Figure 3.4 shows a steady state analysis using $S_0 = 4$ gL^{-1}, the standard sugar concentration in supplemented M9 media, and the dilution rate $D = 0.1$ h^{-1} which is 10% of the value chosen as the maximal growth rate of the killer strain. Steady states are present for AHL production rates of $k_A \leq 10^{-20}$ mol cell^{-1} h^{-1}, though all of the states in which both species present are unstable, and therefore undesirable.

Figure 3.5 shows the steady states of the system over a range of values of S_0, D and k_A. As S_0 is increased, so too does the population level of the competitor strain at steady state. Similarly, as D increases, so too does the population level of the killer strain at steady state. This is intuitive as the richer media, being replenished more frequently, would be expected to support larger populations. An interesting point that arises from this analysis is that the only stable steady states present are at low dilution rates and an AHL production rate of $k_A = 10^{-19}$ mol cell^{-1} h^{-1}. Under these conditions one unstable and one stable foci are present; with a higher killer population at the stable steady state.

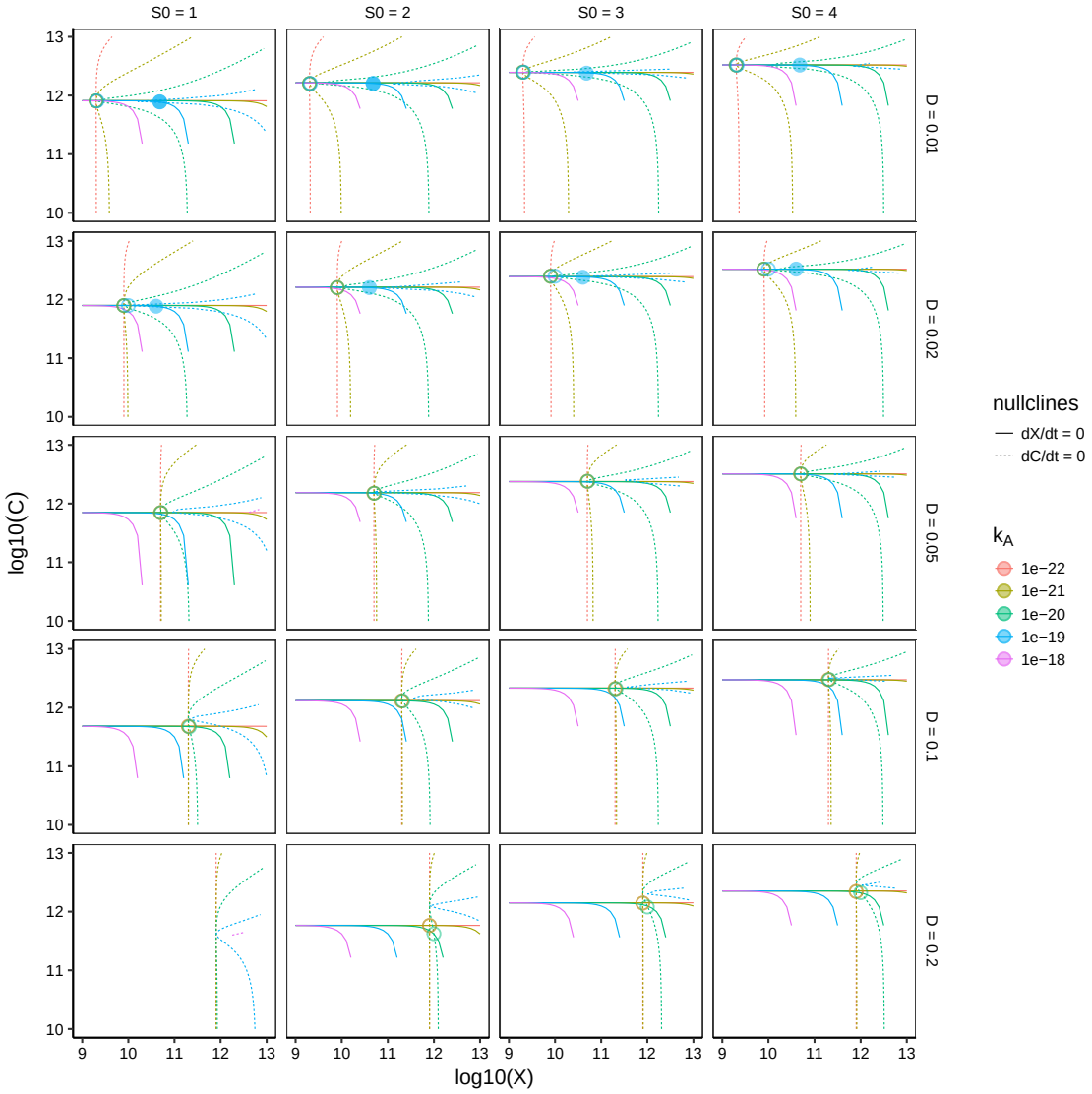

Figure 3.5: Steady state analysis of simple SPoCK in chemostat model over a broad range of system conditions. The different colours show the behaviour with different rates of AHL expression; a parameter controllable in the real system by changing arabinose concentration. The steady state concentrations of each species are shown as circles with empty circles showing saddle nodes and filled circles showing stable nodes or foci. ($\mu_{X_{max}} = 1\ \mathrm{h}^{-1}$, $\mu_{C_{max}} = 1.2\ \mathrm{h}^{-1}$)

The range over which the AHL production rate, k_A, is varied in Figure 3.5 is 10,000 fold, yet the dynamic range of the bacteriocin expression is unlikely to be close to that. Figure 3.6 shows the steady state analysis over a narrower range of k_A in order to capture some more of the bacteriocin expression range. One can see that there are now many more stable steady states present. At lower values of k_A the steady states are at higher killer densities and very slightly lower competitor

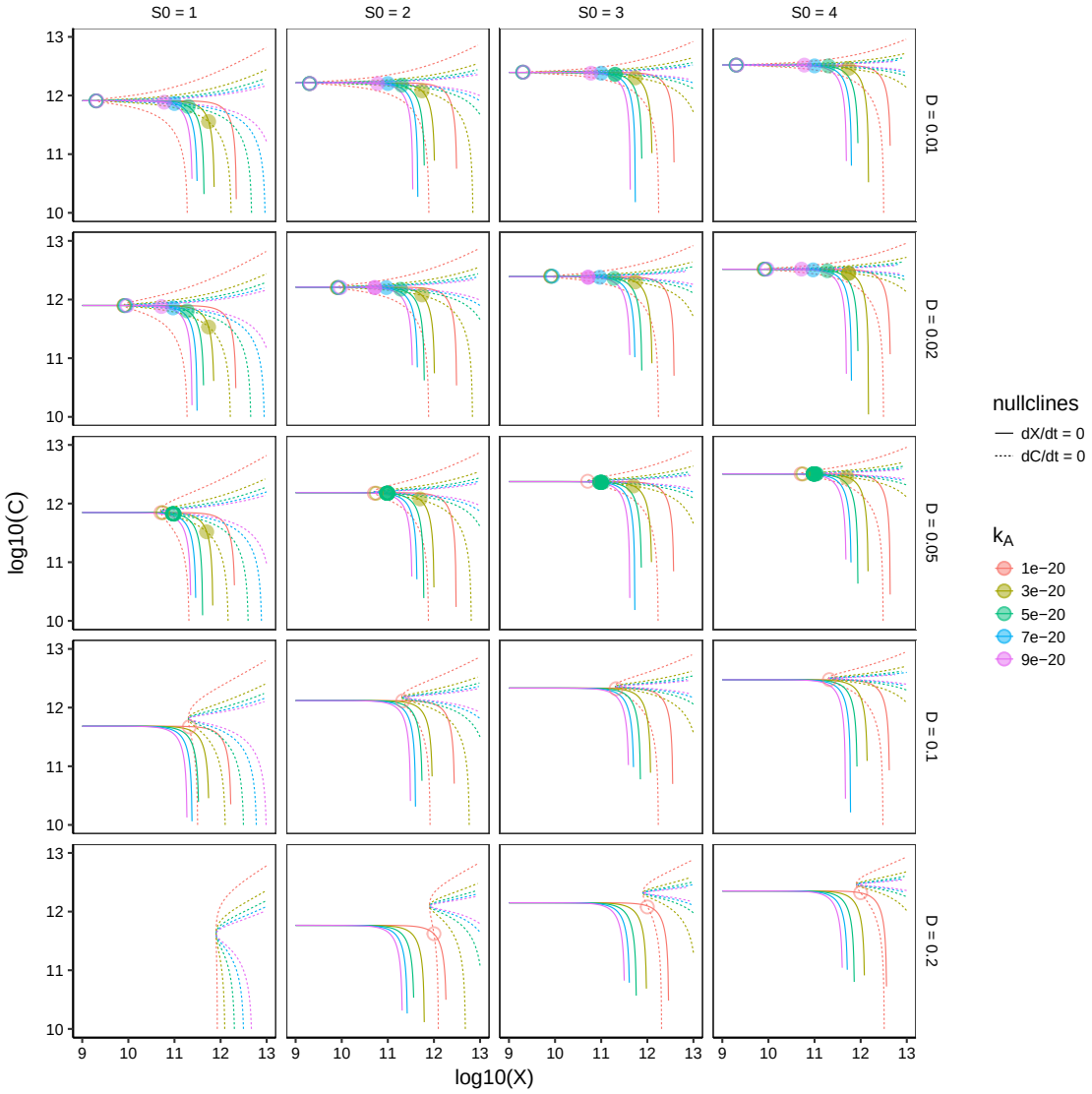

Figure 3.6: Steady state analysis of simple SPoCK in chemostat model over a narrow range of AHL expression conditions. The different colours show the behaviour with different rates of AHL expression; a parameter controllable in the real system by changing arabinose concentration. The steady state concentrations of each species are shown as circles with empty circles showing saddle nodes and filled circles showing stable nodes or foci. ($\mu_{X_{max}} = 1\ \mathrm{h}^{-1}$, $\mu_{C_{max}} = 1.2\ \mathrm{h}^{-1}$)

densities. Further, stable steady states are present at slightly higher dilution rates.

It is also possible to determine whether there are any steady states present at different levels of growth advantage for the competitor strain. Figure 3.7 shows the steady states with varying competitor growth rates over different media and dilution conditions. Interestingly, when the competitor has a higher maximal growth rate, stable nodes are present that do not exist at the lowest competitor growth rate.

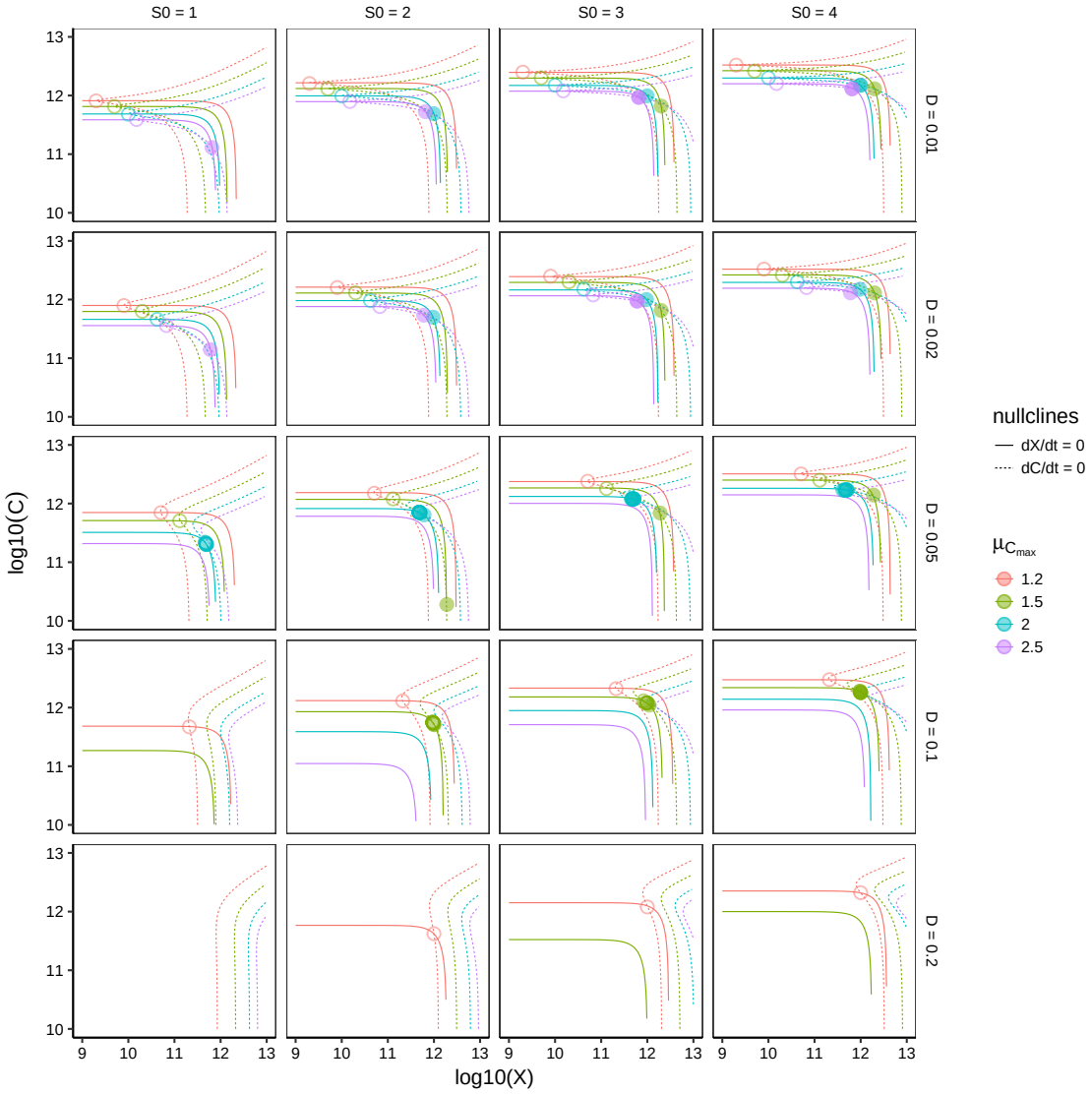

Figure 3.7: Steady state analysis of simple SPoCK in chemostat model with different levels of competitor fitness. The different colours show the behaviour with different maximal growth rates of the competitor strain. The steady state concentrations of each species are shown as circles with empty circles showing saddle nodes and filled circles showing stable nodes or foci. ($\mu_{X_{max}} = 1$ h^{-1}, $k_A = 10^{-20}$ mol cell^{-1} h^{-1})

Further, the position of the steady states moves towards a higher killer and lower competitor population level.

Plotting the trajectories of the killer versus competitor strains shows the progressions towards the stable steady states, Figure 3.8. However, a cluster of stable and unstable steady states identified when $k_A = 5^{-20}$ mol cell^{-1} h^{-1} and $\mu_{C_{max}} = 1.5$ h^{-1} does not seem to act as an attractor; instead the killer population, X, is lost. The

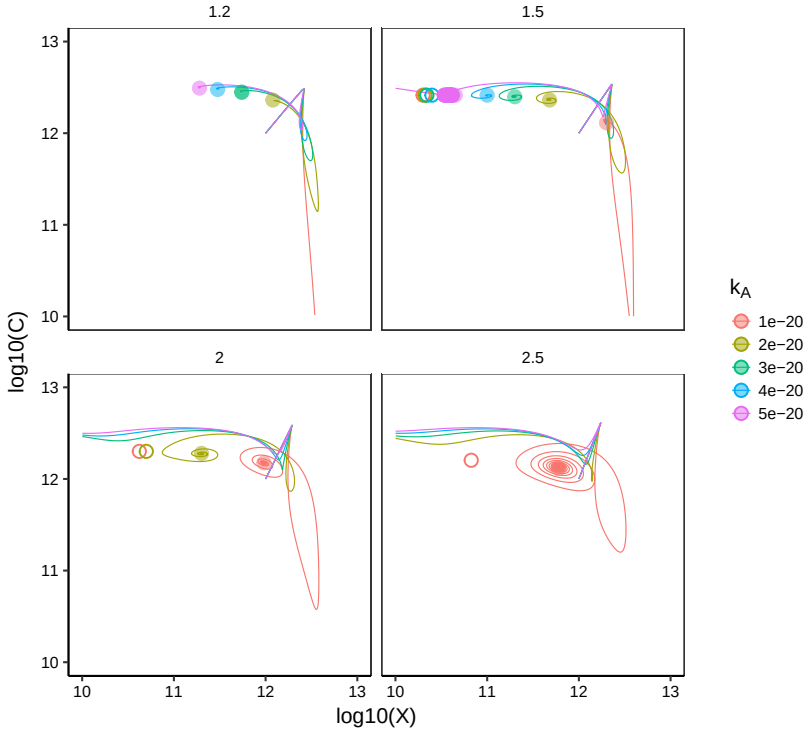

Figure 3.8: Trajectories of the killer and competitor strain in chemostat with varying levels of competitor fitness $\mu_{C_{max}}$ and AHL production k_A. All the trajectories start at $X(0) = C(0) = 10^{12}$ cells L^{-1}. The steady states are shown as circles with empty circles showing unstable nodes and filled circles showing stable foci. ($S_0 = 4$ gL^{-1}, $D = 0.02$ h^{-1}, $\mu_{X_{max}} = 1$ h^{-1})

spirals towards the stable steady states, prominent when $\mu_{C_{max}} \geq 2$ h^{-1}, will appear as damped oscillations over time, as will be shown in the next section. Again, these simulations show that it is possible to produce stable populations of competing cells and tune the relative densities of the two populations using only the rate of AHL production.

3.1.2 Oscillations in the Simple SPoCK Model

Finding an analytical solution to the system of equations making up the SPoCK model is not possible. In addition, techniques for finding limit cycles and other oscillatory states in non-linear systems are either restricted to two species systems or non-trivial. As such, numerical integration of the system over a restricted parameter space is used here to observe conditions under which oscillatory behaviour can be observed. However, in the next Chapter I use statistical methods to examine

oscillations in a more complete model.

Figure 3.9 shows dampened oscillations occurring when the maximal growth rate of the competitor strain is large, $\mu_{C_{max}} \geq 1.6$. The speed of the dynamics is dependent on dilution rate, which is intuitive as in this simple model dilution provides the only means of removal of molecules and species. The maximal bacteriocin production rate and AHL production rate have to be correspondingly increased along with the dilution rate.

3.2 Addition of Self-Regulation

The SPoCK system as described relies on the competitor strain to grow at a faster rate than the killer. In certain industrial applications the "competitor" strain may be producing a recombinant protein that causes a great deal of burden, leading to slow growth. As shown in Figure 3.1F, the addition of a self-regulation mechanism, along with the bacteriocin competitor regulation could allow the SPoCK system to function under a broader range of circumstances.

A simple, abstract addition to the model will suit as a proof of concept of this self-regulation. The killer strain has an extra negative regulation parameter added which is dependent on the AHL concentration, A. This leads to a change to Equation 3.1:

$$\frac{dX}{dt} = (\mu_X(t) - D - \omega_X A(t))X(t) \tag{3.21}$$

where ω_X is the rate of killing of the killer strain due to an AHL induced toxin. A more complete way to model this would be to add a new intracellular toxin species to the set of equations and model it's change due to AHL induction. Figure 3.10 shows how the addition of this self killing leads to a population cap on X which is dependent the rate of AHL production, k_A, and the rate of AHL induced killing. As the rate of AHL production decreases, the set-point for a stable killer population increases. Similarly, as the rate of AHL induced killing increases, the stable set-point decreases.

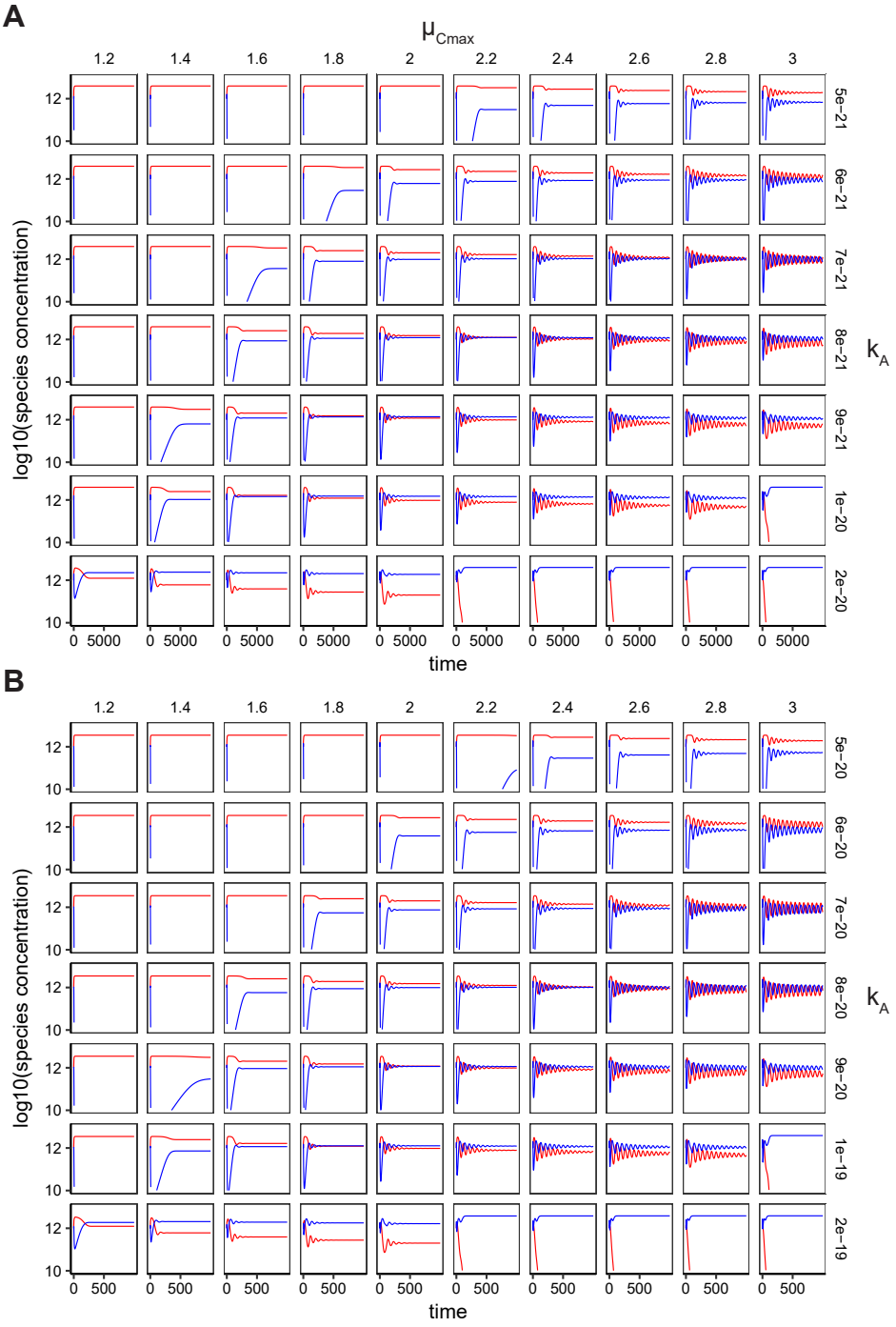

Figure 3.9: Discovering oscillations of the killer (red line) and competitor (blue line) over time across different values of maximal competitor growth rate, $\mu_{C_{max}}$, and AHL production rate, k_A. (A) Low dilution rate, $D = 0.02$ h^{-1}, and slow maximal bacteriocin production rate, $k_{\beta_{max}} = 10^{-22}$ mol cell^{-1} h^{-1}, lead to slow population dynamics and a necessity for lower AHL production rates. (B) Faster dilution rate, $D = 0.2$ h^{-1}, requires a faster bacteriocin production rate, $k_{\beta_{max}} = 10^{-20}$, and faster AHL production to achieve the same behaviour.

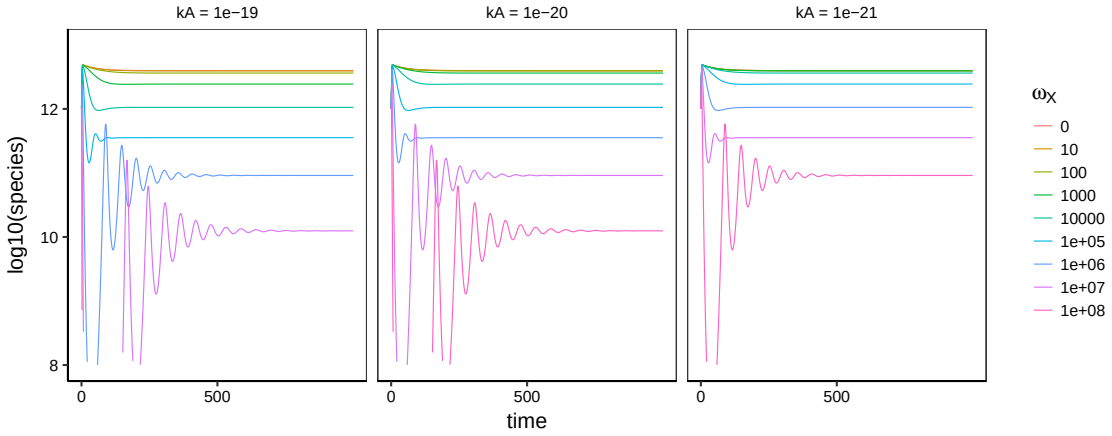

Figure 3.10: SPoCK with self-regulation. With no competitor present, the self-regulation mechanism can produce different stable levels of killer bacteria. The coloured lines show different rates of self-killing, ω_X.

When the model is run with competition the self-regulation mechanism is able to create several stable population levels, including when the competitor is slower growing. Such simulations can be seen in Figure 3.11.

3.3 Discussion

In this Chapter I have described the need for systems that can overcome competitive exclusion to allow heterogeneous communities of bacteria to stably coexist. I have shown that current approaches involve the engineering of all species within the environment, which in some circumstances can be undesirable or impossible. Using bacteriocins to target competing species within an environment, and relieve the exclusionary pressure, is one possible approach to producing a system that enables stable communities while only necessitating the engineering of a single strain. The use of quorum sensing as a modulator on the production of the bacteriocin and the inherent fitness advantage of the non-engineered competitor prevents the engineered strain from overtaking.

Using a simple mechanistic model of the SPoCK system in a chemostat environment, I have shown that the system can stabilise two populations. Stable steady states can be found over a wide range of realistic environmental conditions. Importantly, parameters that are theoretically easy to control in an experiment can be

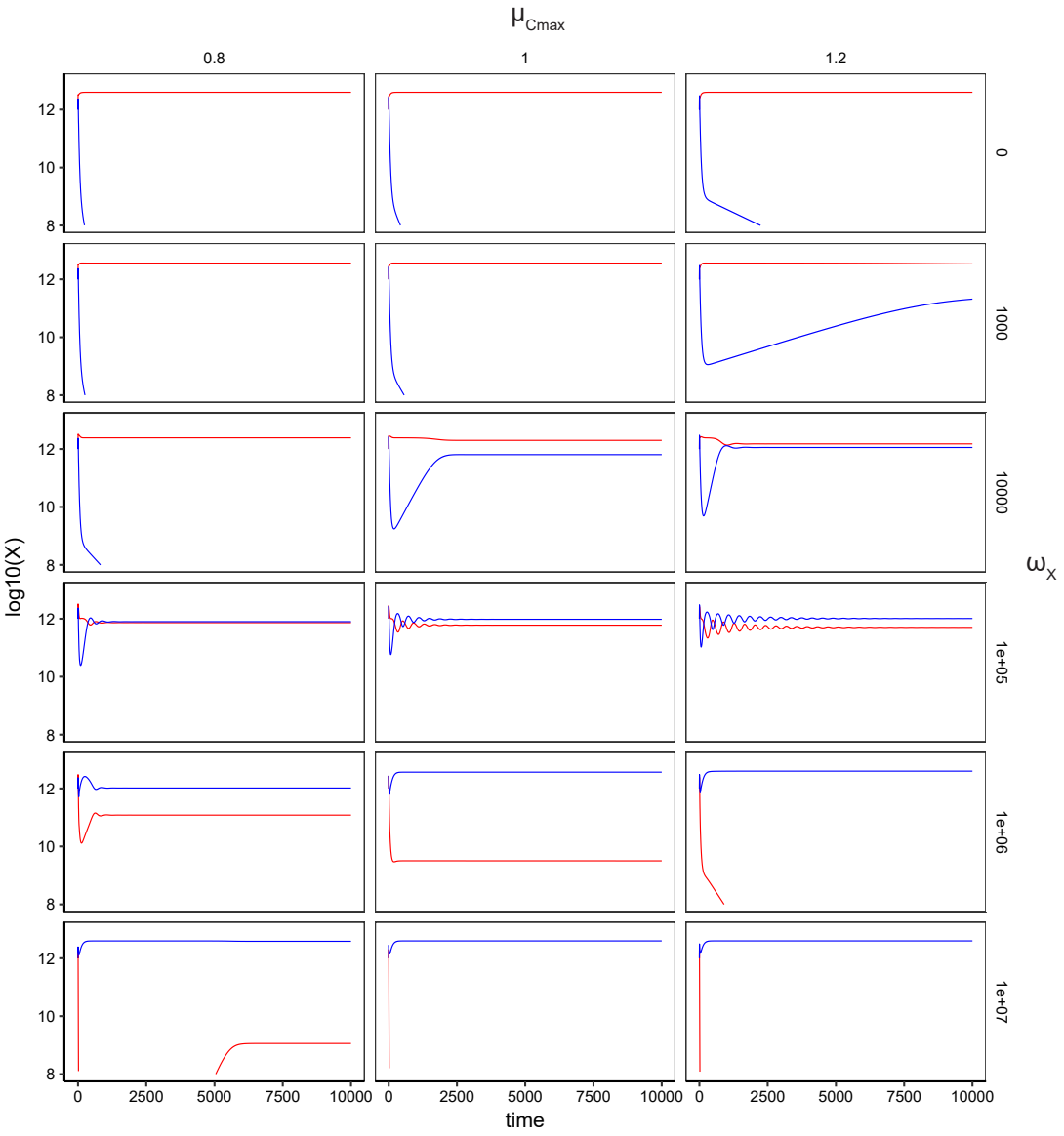

Figure 3.11: SPoCK with self-regulation in competition. The red line shows the killer strain and the blue line shows the competitor.

used to tune the population levels of the two strains. Finally, oscillatory states with different periods have been identified. The period is dependent on the dilution rate, with faster dilution resulting in faster turnover of cells, AHL and bacteriocin in the environment. However, this must be correspondingly matched with increased production rates of AHL and bacteriocin and competitor growth rates.

The complexity of the model, even this simplified one, limits the types of analyses that can be used to explore its dynamics. As extra layers of complexity are

added, as in Figure 3.1, the complexity of the model will duly increase.

A model was produced showing a simple addition of a self-regulation mechanism in to the killer strain. Simulations with this model showed that changes in AHL production and killing rate could produce stable population over a wide range of levels. When in competition, the self-regulation allowed the stable maintenance of two populations, even when the non-killing competitor grew more slowly. Although this model was a simple and abstract extension of the original SPoCK chemostat model, it does show that additional feedback systems can be used to create a variety of different behaviours.

In the next Chapter I will show the construction and characterisation of the SPoCK system in *E. coli*. A more realistic, and complex, model is built which will rely on Approximate Bayesian Computation (ABC) to predict the system's dynamics.

3.4 Conclusion

In conclusion, I have described how the construction of planktonic, multi-species populations in synthetic biology are becoming increasingly necessary. However, I have detailed how, currently, the design of such system is not possible due to competitive exclusion. I have designed and modelled the first system to use dynamic production of bacteriocins to counteract competitive exclusion. This is also the first system in which control of multiple species within a community is possible through the engineering of a single constituent strain. The modelling demonstrated that this system is able to produce stable equilibria or oscillating populations, tunable via easily controllable parameters. Finally, I demonstrate how such a system could be made more robust through the addition of a negative feedback loop on the controller population.

3.5 Future Work

Several models for population control were suggested at the beginning of this Chapter. Exploration of the behaviours that these systems are capable of producing may reveal that one of these models can more robustly produce a desired set of be-

haviours. This can be done using approximate Bayesian computation [111]. Many of the systems suggested can reuse modules from the SPoCK system.

Chapter 4

Construction and Characterisation of Stabilised Populations by Community Killing (SPoCK)

In the previous Chapter I detailed a theoretical system for the stabilisation of a community of competing bacterial strains. The system involved the density dependent production of a bacteriocin by a slow growing strain to inhibit the growth of a fitter competitor. In this Chapter I will describe the construction and characterisation of such a system.

4.1 Controllable Bacteriocin Production

The microcin-V system, that was used in Chapter 5, provides a simple bacteriocin system to work with; comprised of just four genes and no post-translational modification of the bacteriocin peptide [119]. This simplicity has meant microcin-V has been used to explore the hybridisation of bacteriocins from different sources to increase the spectrum of activity [142, 143]. It has also been used as the basis of a modular bacteriocin delivery system due to the transport mechanisms requiring a simple glycine-glycine signal peptide for secretion [120]. There is a large amount of literature on how bacteriocins are induced in competitive environments [144, 145, 146], but, to my knowledge, there is no literature on the creation of an inducible or repressible bacteriocin system.

In this work the PLtetO-1 promoter is used in place of the native putative promoter upstream of the cvi immunity and cvaC bacteriocin genes. The PLtetO-1 promoter was chosen as it provides very low leakiness in its "off" state and a large regulatory range [147]. A system in which the expression of the microcin-V bacteriocin gene, cvaC, is repressible by TetR was constructed, Figure 4.1a. The expression of the transport proteins, CvaA and CvaB, were kept under the control of the native promoter. The expression of the immunity, cvi, and bacteriocin, cvaC genes was driven by the PLtetO-1 promoter, with the native putative RBSs of both genes maintained. The plasmid expressed tetR from a Ptrc promoter which is repressible by LacI and derepressed through the addition of IPTG. Additionally, GFPmut3b is expressed from a second PLtetO-1 promoter as a proxy for bacteriocin expression. The dynamics of the system were explored using flow cytometry to measure GFP levels 24 hours after induction with varying levels of aTc, Figure 4.1b. Low levels of induction occur at 10^{-8} g/mL of aTc and increase with increasing inducer concentration. To determine whether bacteriocin was being produced, the induced culture was spotted on a lawn of *E. coli* MG1655 and incubated overnight, Figure 4.1c. This shows that even at very low concentrations of inducer, some inhibition of growth occurs. This inhibition increases as aTc concentration increases.

Analysis of the native 5' upstream region of the cvi and cvaC genes using promoter identification software BPROM [148] indicates that, in addition to the promoter upstream of the cvi start codon, there may be a promoter region within the cvi gene upstream of the cvaC start codon, Figure 4.2. This means that in the TetR repressible system shown in Figure 4.1a, complete inhibition of the PLtetO-1 promoter may not be enough to prevent expression of the bacteriocin.

In light of this finding, a new TetR repressible bacteriocin plasmid was constructed based on the pMPES:V plasmid [120]. In this system, the start codon of cvaC in the native microcin-V cassette is mutated to prevent expression of the bacteriocin from the cassette but allowing native expression of the transport proteins and immunity. Separate cvaC and cvi genes, with non-native RBSs, were synthesised and placed under the control of the ProTeOn expression system [149]. In this

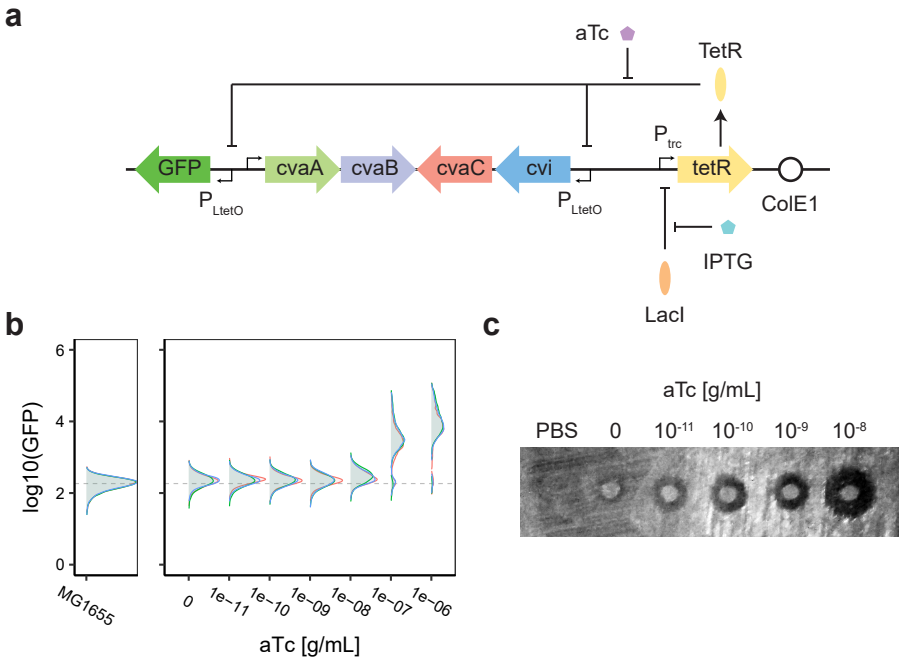

Figure 4.1: Inducible expression of microcin-V. a) A schematic of the inducible bacteriocin system. The expression of the bacteriocin, cvaC, and immunity, cvi, are from the PLtetO-1 promoter. This is repressed by TetR which is expressed from a Ptrc promoter. aTc can be added to dis-inhibit expression of the bacteriocin. IPTG is added to ensure that expression of TetR is maximal. Production of the transport proteins, cvaA and cvaB, are driven by the native promoter. GFP-mut3b is also expressed from a PLtetO-1 promoter as a proxy for bacteriocin expression. b) Flow cytometry data showing GFP expression after 24 hours after induction with different concentrations of aTc. c) Inhibition zones on a lawn of *E. coli* MG1655. Bacteriocin producing cultures were induced for 24 hours with varying levels of aTc and then spotted on to the lawn and incubated overnight.

work, this plasmid was altered by replacing the ProTeOn expression system with a PLtetO-1 promoter and GFPmut3b gene upstream of the functional cvaC gene. This plasmid does not carry it's own mechanism for the production of TetR. The characterisation of this system will be described in Section 4.3.

4.2 A Tunable Quorum Sensing System

The SPoCK system requires a quorum signal to produce the TetR protein which represses the production of the bacteriocin, as shown above. There are many quorum sensing systems found in nature that have been adapted for use in synthetic circuit

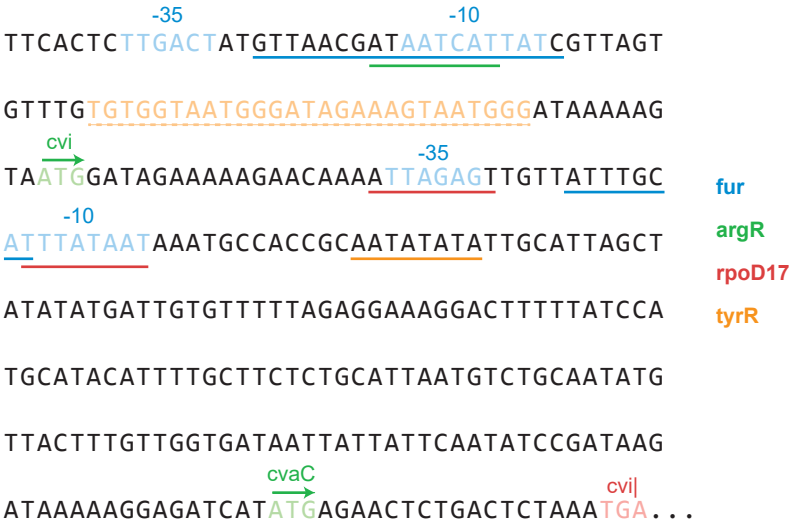

Figure 4.2: The cvi and cvaC native promoter region. Putative -35, -10, translation start sites and transcription factor binding sites identified using BPROM. -35 and -10 sites are highlighted with light blue font. Translation start sites for cvi and cvaC are highlighted in green font. The translation stop codon for the cvi gene is highlighted in red font. The transcription factor binding sites for fur (blue), argR (green), rpoD17 (red) and tyrR (orange) are underlined with solid lines. The primer binding site used for the construction of plasmid pAF006 is highlighted in orange font and underlined with an orange dashed line.

engineering. The important issues to consider when choosing which system to use are simplicity/complexity and orthogonality. The simplest systems require a protein to produce diffusible quorum molecules and a transcription factor which, when bound to those molecules, controls the activity of a promoter [150]. More complex systems utilise multiple membrane bound receptors which are able to integrate multiple signals, or produce complex signalling cascades [151]. Orthogonality concerns impact whether other cellular mechanisms within a chosen host are regulated by quorum sensing molecules or the host produces quorum molecules that may interfere with predictable circuit behaviour. Further orthogonality concerns may be present when choosing multiple quorum sensing systems to use in a circuit or group of circuits [152].

Here the well understood luxR-luxI system is used. The wild-type *lux* system incorporates a bi-directional promoter producing LuxI, the quorum molecule synthase, in the forward direction and LuxR, the transcription factor, in the reverse direction as shown in Figure 4.3a. LuxI produces the quorum molecule

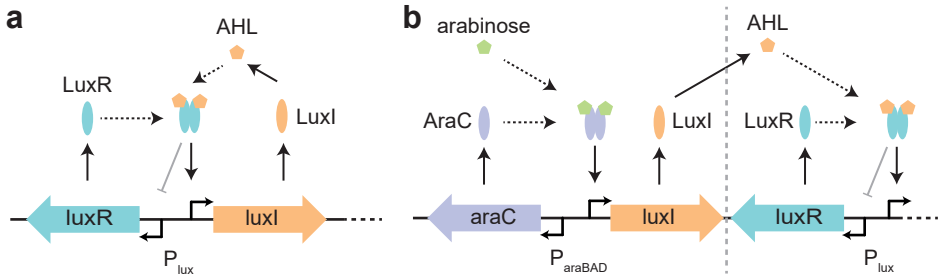

Figure 4.3: Design of an inducible quorum sensing system. a) The wild-type *lux* system. The PLux promoter transcribes the quorum molecule synthase, LuxI, in the forward direction and the transcription factor, LuxR, in the reverse direction. The quorum molecule binds to LuxR which activates the forward transcription of the PLux promoter. b) An arabinose inducible quorum system. Two distinct parts of the system are separated by the grey dashed line; the inducible quorum molecule production and the sensing of the quorum molecule.

3-oxohexanoyl-homoserine lactone (3OC6HSL), abbreviated here to AHL, which freely diffuses in and out of the producing cells. The transcription factor, LuxR, forms a complex with the AHL and dimerises with another ligand bound LuxR to form the active transcription factor. This active transcription factor dimer binds to a region within the *lux* promoter, known as the *lux*-box, and greatly increases forward transcription. The LuxR-AHL complex also represses transcription in the reverse direction [153].

In order to make this system tunable the production of LuxI is placed under the control of the arabinose inducible promoter ParaBAD [154]. Previously designed inducible quorum sensing systems have also placed the production of LuxR under a controllable promoter [155]. However, here LuxR is left under the control of the PLux promoter as in [156], as shown in Figure 4.3b.

In order to characterise the dynamics of this inducible quorum sensing system, the circuit was constructed with fluorescent proteins as readouts from the two forward promoters. The circuit was constructed in two steps; the arabinose inducible *luxI* section and the *lux* promoter section were first built on separate plasmids, after which the sections were combined onto a single plasmid. The cloning process is described in detail in Section 2.2.

4.2.1 Lux Promoter Dynamics

The response of the *lux* promoter to AHL is examined in Figure 4.4. The fluorescence response after 20 hours can be seen in the flow cytometry data in Figure 4.4b. There is a $\sim$ 1000 fold increase from basal expression to maximal expression. The data also show that rather than a homogeneous population response of increasing expression with increasing inducer concentration, there is a binary cellular response of expression being on or off and a threshold concentration, at 10nM AHL, at which the we can observe a heterogeneous population. The growth data in Figure 4.4c shows that at higher levels of induction growth may be slightly impacted. The fluorescence time-course in Figure 4.4d and the per cell fluorescence time-course in Figure 4.4e demonstrate that inducer concentrations of 100nM and above saturate the system, leading to identical dynamics. There is a dip in fluorescence in all populations showing expression after $\sim$ 7 hours. This indicates that expression rate from the lux promoter has dropped below the degradation rate of the sfGFP-LAA. This may be due to growth phase dependant repression of the promoter by one of the LysR-family transcriptional regulators [157] or a lack of resource availability such as RNAP or ribosomes [158].

4.2.2 araBAD Promoter Dynamics

The second constituent part of the tunable quorum sensing system is the inducible expression of LuxI. Here the inducible promoter used is the ParaBAD system which takes arabinose as an input to de-repress the promoter. To determine dynamics of expression from the ParaBAD system, a blue fluorescent protein (mTagBFP2) was used, with *luxI* with an LAA-ssrA degradation tag cloned downstream, as shown in Figure 4.5a. Expression of BFP at various induction levels was measured using flow cytometry, Figure 4.5b. This shows a system inducible over a $\sim$ 1000 fold dynamic range, though with an expression increase of less than 100 fold. The expression at the maximum inducer concentration used here may not be the maximum achievable expression level; the plot doesn't show the expression plateauing. However, as shown in Figure 4.5c, at the highest level of inducer concentration used here, the growth dynamics change. This may be due to the bacteria using the arabinose

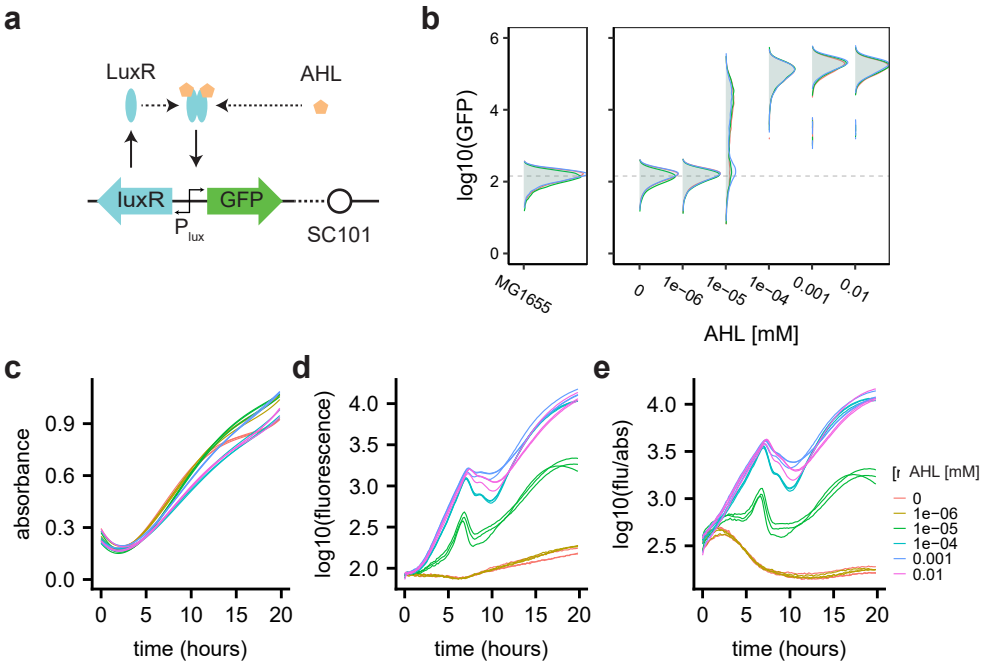

Figure 4.4: The dynamics of the *lux* promoter. a) Forward transcription from the *lux* promoter, induced by the addition of AHL to the growth media, expresses superfolder GFP with the LAA-ssrA degradation tag. b) Flow cytometry data showing the expression of GFP following 20 hours of induction with different concentrations of AHL. The red, green and blue lines show are three separate replicates at each concentration. The grey dashed line shows the median fluorescence of MG1655 with no plasmid. c, d, e) Time-course of induction at various concentrations of AHL. Three replicates for each inducer concentration. The lines are Loess curves through data sampled every 10 minutes. c) Absorbance at 540nm. d) log10 fluorescence measurement with excitation at 480nm and emission 510nm. e) log10 fluorescence per unit absorbance as a proxy of GFP per cell.

as a carbon source. Due to equipment limitations, it was not possible to measure BFP expression concurrently with absorbance. Instead the bacteria were induced with 1.33mM arabinose in phosphate buffered saline (PBS), as in [159], at different population densities, Figure 4.5d. The BFP fluorescence was measured using flow cytometry after 6 hours, Figure 4.5e. The results show comparable BFP expression across all dilution levels demonstrating expression is independent of population density.

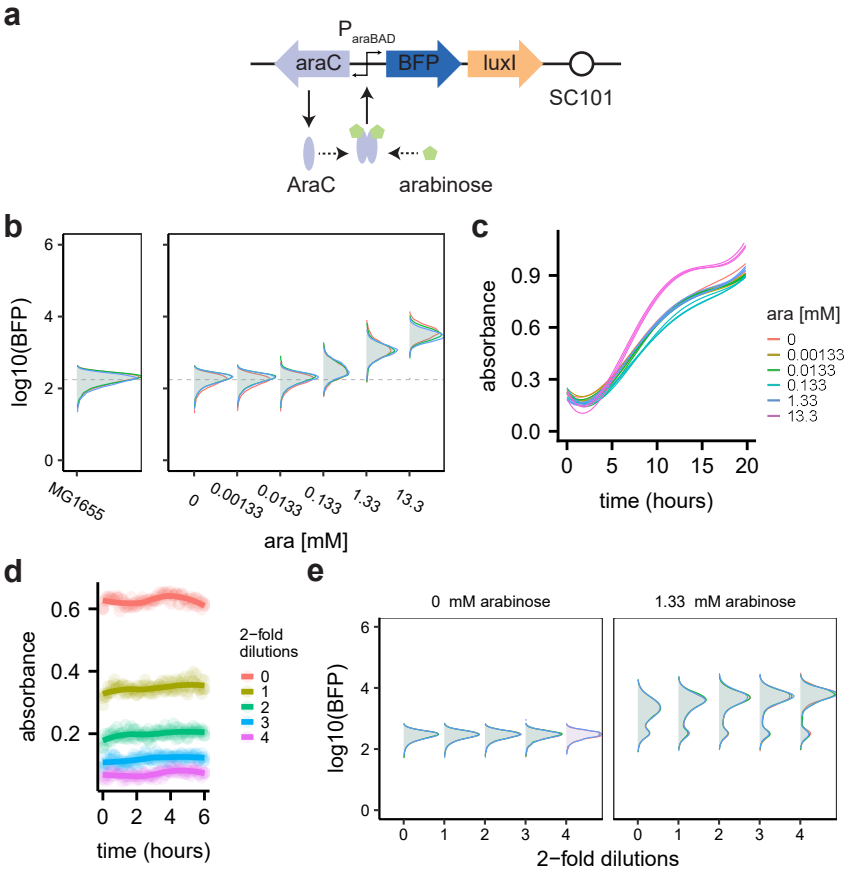

Figure 4.5: The dynamics of the ParaBAD promoter. a) Forward transcription from the ParaBAD promoter, induced by the addition of arabinose to the growth media, expresses *mTagBFP2* and *luxI*. b) Flow cytometry data showing the expression of BFP following 20 hours of induction with different concentrations of arabinose. The red, green and blue lines show are three separate replicates at each concentration. c) Time-course of growth at various induction concentrations of arabinose. Three replicates for each inducer concentration. Absorbance at 540nm. Due to instrument limitations it was not possible to measure BFP fluorescence during the time-course Instead various dilutions of cells in phosphate buffered saline (PBS) were induced with 1.33 mM arabinose. d) Cell density is maintained over 6 hours of induction in PBS. e) Flow cytometry measurements of BFP expression after 6 hours of induction in PBS show that expression is density independent.

4.2.3 Dynamics of the Tunable Quorum Sensing System

The two parts of the system, described above, were cloned into a plasmid with an SC101 origin of replication and kanamycin resistance. This was transformed into *E coli* MG1655. The dynamics of the system were determined through induction with different concentration of arabinose. Figure 4.6a shows how introduction of

arabinose leads to expression of BFP and GFP through the activation of the two different promoters. The BFP expression, Figure 4.6b, shows the same dynamics as described above. GFP expression shows the same minimal and maximal induction levels as with induction by AHL, Figure 4.6c, indicating that the arabinose inducible part of the system is capable of triggering the production of enough AHL to saturate the lux promoter, even with the LAA-ssrA degradation tag on the LuxI protein. The correlation between BFP and GFP expression, Figure 4.6d, shows that the threshold for maximal expression of GFP from the lux promoter requires a very small increase from basal levels in expression of BFP from the ParaBAD promoter. However, using the Operon Calculator (v1.0) [160] the translation rate of LuxI is ~ 10 fold greater than that of BFP in this system. The growth time-course, Figure 4.6e, shows again that at the highest concentration of arabinose there is increased growth but at all other concentrations growth dynamics are comparable. The population level fluorescence time-course, Figure 4.6f, and "per cell" fluorescence, Figure 4.6g, show some unexpected dynamics. At 0.133mM, the lowest concentration to illicit a BFP or GFP response, the expression of GFP occurs early followed by a dip similar to that seen in Figure 4.4d. After this dip, expression of GFP increases at a slower rate than for higher concentrations of arabinose induction, as one would expect. At the highest level of induction, 13.3mM arabinose, production of GFP matches the increase in population density until the cells begin to reach stationary phase, as can be seen by the flat period from the 5 hour mark in Figure 4.6g. At 1.33mM arabinose induction, there is a similar but shorter period of GFP production matching growth rate. However, the expression rate increases sooner than for the higher inducer concentration. This may be due to the slightly lower growth rate leading to greater availability of cellular resources.

4.2.4 Plasmid Copy Number Effects on Quorum Sensing System

Experiments up to this point have been carried out using a plasmid with an SC101 low copy origin of replication. The only lever for control for this system is the concentration of arabinose. In order to gain access to a wider landscape of quorum sensing dynamics, the plasmid copy number was varied. It has been shown that copy

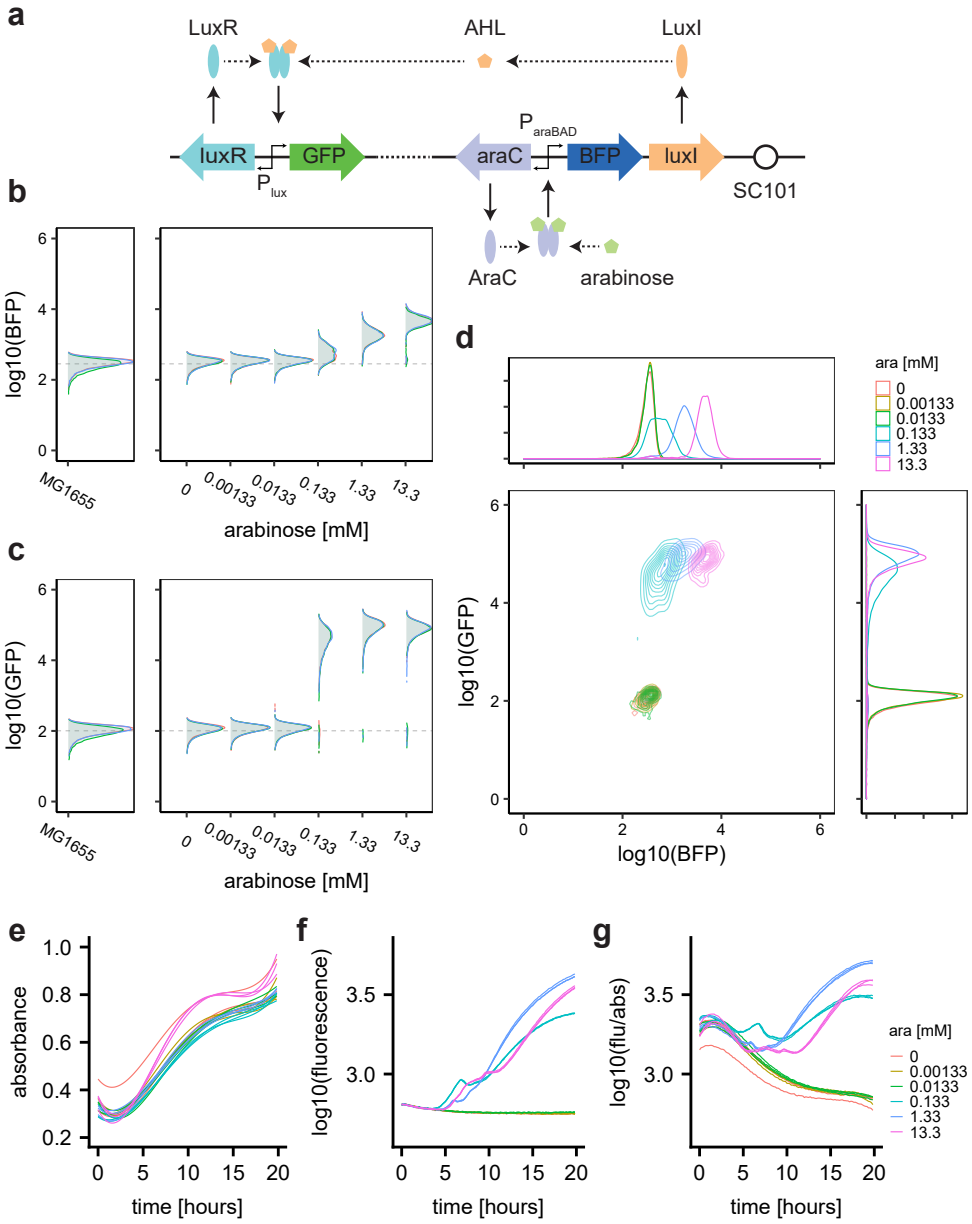

Figure 4.6: a) The full tunable quorum sensing circuit. b, c & d) The BFP and GFP fluorescence measured by flow cytometry after 20 hours of induction with varying levels of arabinose. e, f & g) Absorbance and fluorescence time-courses with induction at various concentrations of arabinose.

number is an important parameter in the functioning of synthetic biological circuits [22] though this doesn't make available the continuous scale of tunability that the inducible promoter offers. Here, in addition to the low copy SC101, the BBR1 and BR322 medium copy and pUC high copy origins of replication are tested, as show in Figure 4.7a. The high copy nature of the pUC based plasmid and the leakiness

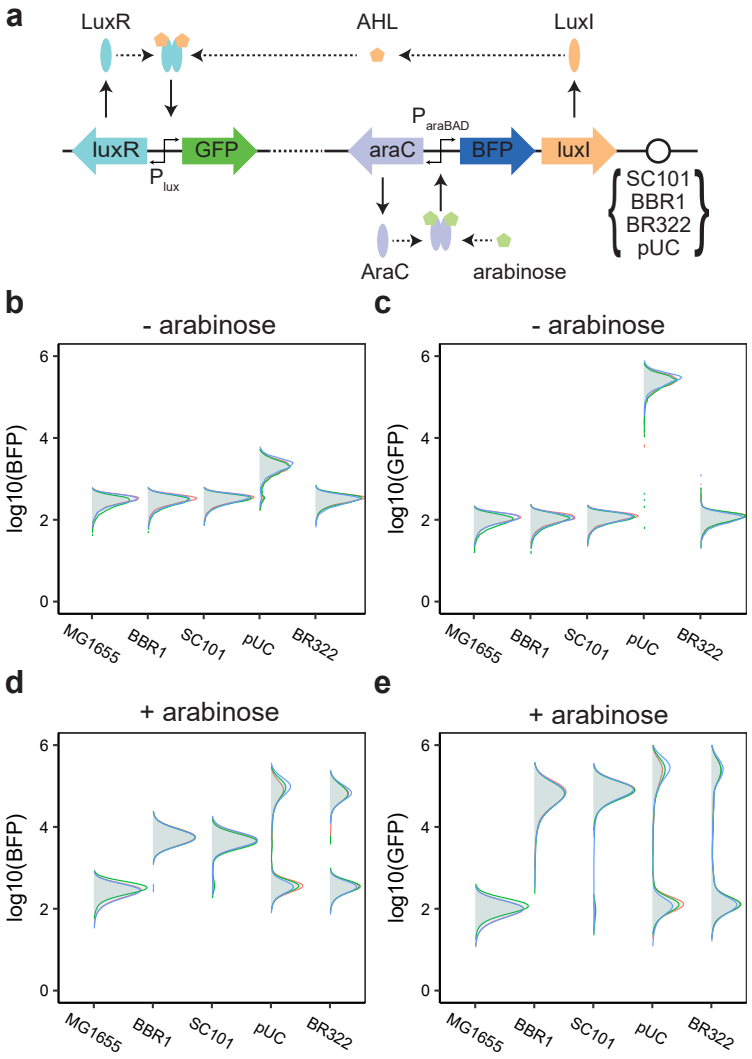

Figure 4.7: Copy number changes to tunable quorum sensing system. a) A cartoon of the plasmid with different origins of replication. b & c) BFP and GFP expression after 20 hours of growth with no induction. d & e) BFP and GFP expression after 20 hours of growth with 1.33mM of arabinose.

of the two promoters leads to the expression of BFP and GFP even without any arabinose induction, Figure 4.7 b & c. The lower copy number plasmids remain in an off state when uninduced, though some BR322 based systems show low levels of GFP expression. When induced with 1.33mM of arabinose, as shown in Figure 4.7 d & e, the BBR1 based system behaves in much the same way as the SC101 based system, though with fewer cells expressing fluorescence of BFP and GFP at low levels.

The BR322 and pUC based systems produced heterogeneous populations with

some cells expressing BFP and GFP to very high levels and some not expressing either fluorescent protein. This indicates that the burden caused by induction of the system has lead to the mutation of one or both parts of the quorum sensing system. Figure 4.8 shows the failure of the BR322 based system with increasing levels of arabinose induction. At the lowest inducer concentration the system functions as expected, with BFP and GFP both being expressed. However, with increasing arabinose concentrations, a sub-population expressing neither BFP nor GFP arises. This indicates that the ParaBAD part is malfunctioning, preventing the expression of BFP and LuxI. The lux part of the system must also be malfunctioning since there is still a population that is producing, secreting and responding to AHL. At higher levels of induction a third population arises, producing BFP but no GFP.

4.3 The Complete SPoCK System

The complete SPoCK system consists of a plasmid carrying the TetR repressible bacteriocin, as detailed in Section 4.1, and a plasmid carrying the tunable quorum sensing system, detailed in Section 4.2. Both plasmids were transformed into *E. coli* MG1655, as shown in Figure 4.9, and the dynamics were explored with a range of origins of replication on the quorum sensing plasmid.

Initially the system was tested with AHL induction to determine that TetR was being produced at a level that could repress the production of the bacteriocin, as measured through GFP expression, Figure 4.10. Although the arabinose inducible part of the quorum sensing system is present on the plasmid, induction with AHL bypasses it, Figure 4.10a. Figure 4.10b confirms using flow cytometry that the arabinose promoter is not induced. The expectation is that GFP expression is high without AHL present and decreases as more AHL is added. This is shown in the flow cytometry data in Figure 4.10c. At the highest inducer concentration, GFP levels are still higher than the control *E. coli* MG1655. This may be due to the long time required for degradation of the GFPmut3b [161] leading to the expression measurements reflecting earlier expression levels.

Next, the SPoCK system is tested using arabinose induction. As with the char-

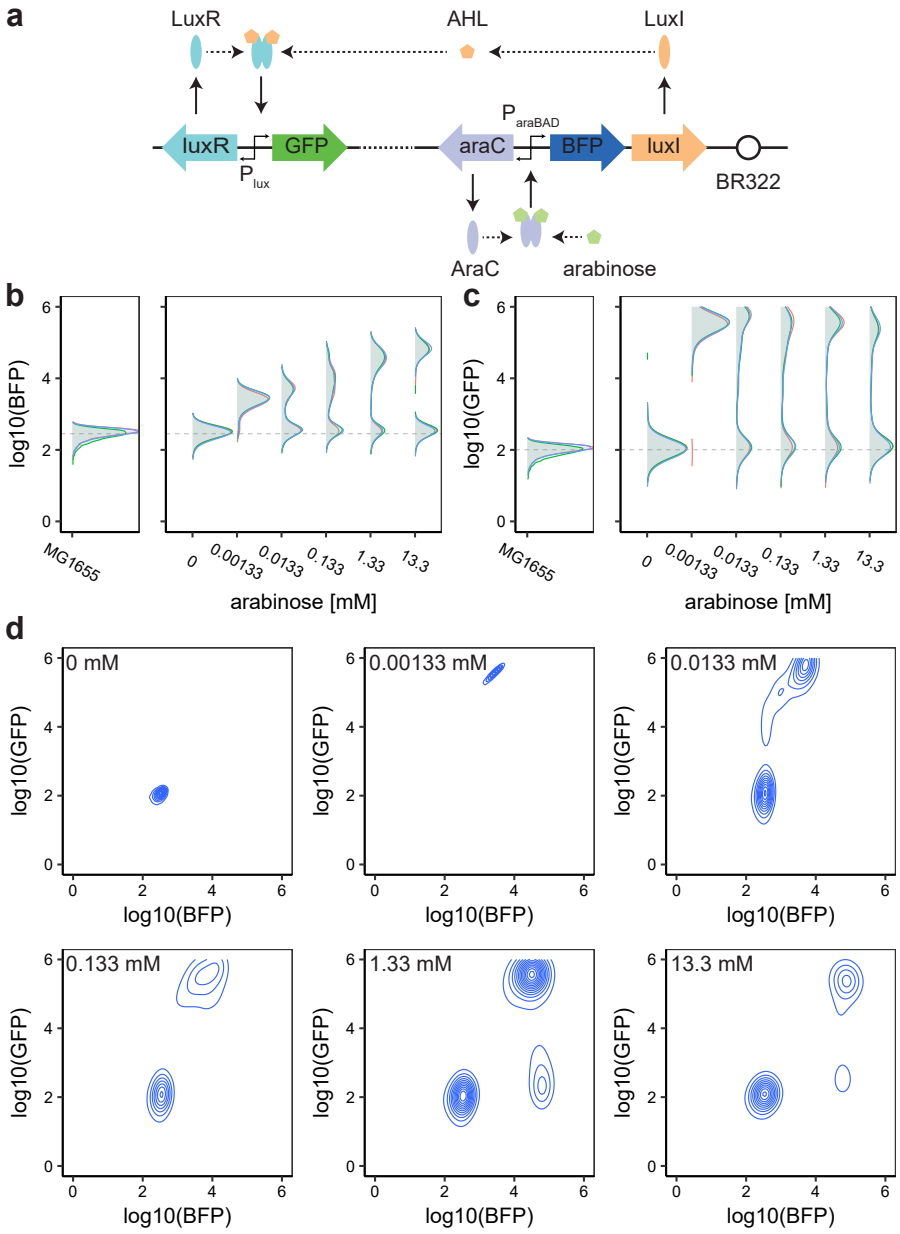

Figure 4.8: Failure of BR322 based quorum sensing system. a) A schematic of the quorum sensing plasmid. b & c) The BFP and GFP expression measured by flow cytometry 20 hours after induction with arabinose. d) The co-expression of BFP and GFP 20 hours after induction with arabinose.

acterisation of the quorum sensing system, arabinose induces luxI expression with mTagBFP2 as a measurable proxy. The LuxI protein is a synthase that produces the diffusible AHL molecule that leads to activation of the lux promoter. The lux promoter drives expression of tetR which inhibits the PLtetO-1 promoter. GFPmut3b is used as a measurable proxy for expression from the PLtetO-1 promoter, which also

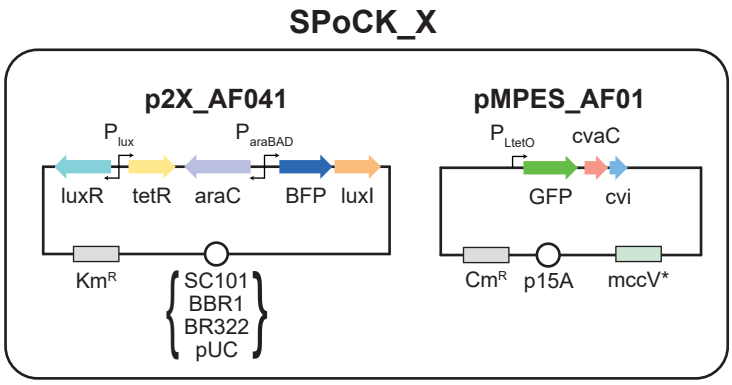

Figure 4.9: The two plasmids of the SPoCK system. The quorum sensing plasmid p2X_AF041 was tested with several origins of replication providing a range of copy numbers. For each origin of replication the 'X' in the plasmid name is replaced with its corresponding SEVA identifier i.e. the SC101 based plasmid is p27_AF041. The killing plasmid, pMPES_AF01, carries the mccV cassette with a mutated cvaC start codon, preventing the production of the bacteriocin from the cassette. Instead bacteriocin is produced from an operon driven by the PLteO-1 promoter.

drives the expression of the bacteriocin, cvaC, and immunity, cvi, Figure 4.11. BFP expression dynamics, Figure 4.11b, are comparable to those measured in the characterisation of the quorum sensing system. Expression of GFP from the PLtetO-1 promoter, Figure 4.11c, is high when there is no arabinose induction, as expected and comparable to the results in Figure 4.10c. However, GFP expression is not inhibited to the same extent at higher levels of arabinose induction. The growth curve time-courses, Figure 4.11d, show comparable results to the arabinose inductions of the quorum sensing system; the highest concentration of arabinose leads to greater growth but all other concentrations show similar dynamics. The fluorescence expression timecourse, Figure 4.11e, and fluorescence "per cell", Figure 4.11f, show a similar binary behaviour to that seen with AHL induction of the SPoCK system in Figure 4.10e. However, the arabinose induction leads to a less pronounced difference between the high and low induced populations. Further, the replicates at the highest arabinose concentration have the highest total fluorescence levels throughout the time-course. This is only slightly offset when divided by the absorbance to give an estimate of per cell fluorescence. The flow cytometry data, Figure 4.11c, does show that the highest arabinose concentration leads to the greatest repression

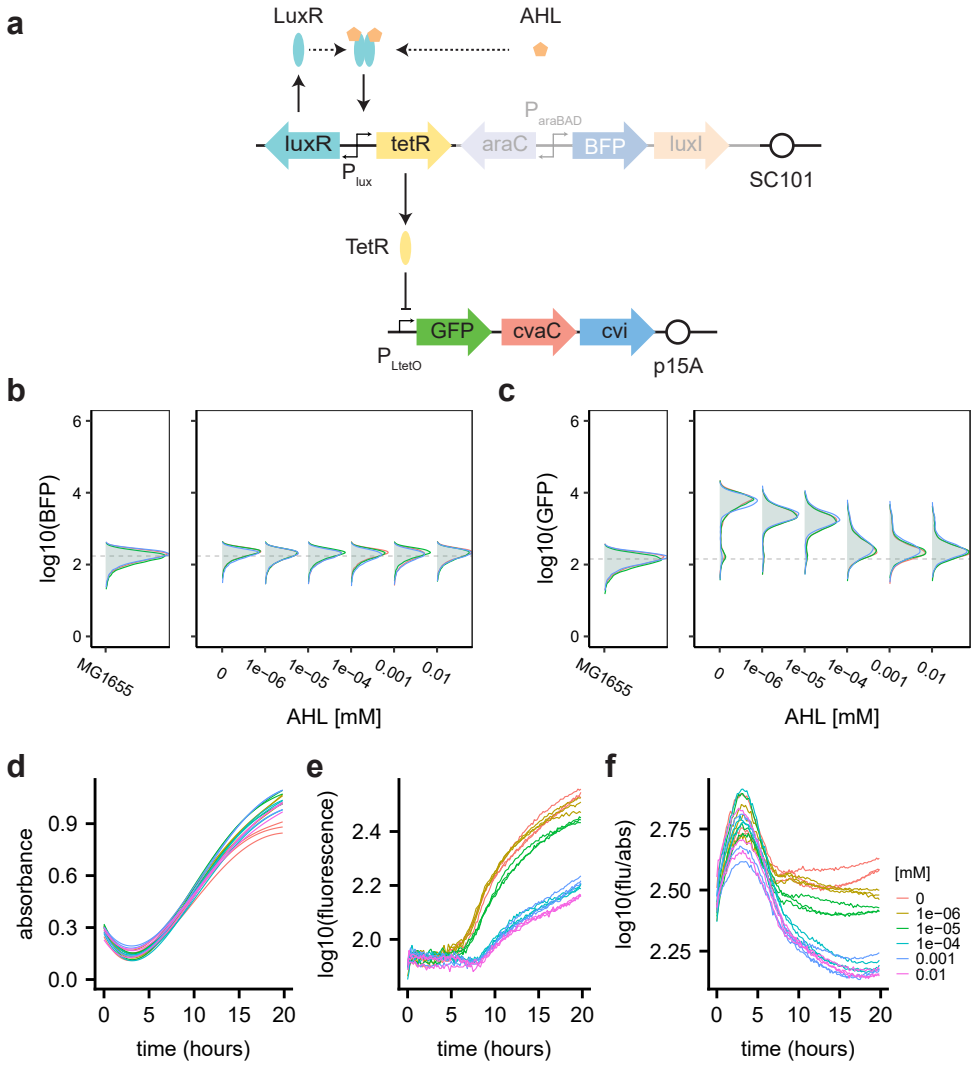

Figure 4.10: Inhibition of bacteriocin production in the SPoCK system with AHL. a) A cartoon illustrating the mechanism of inhibition of bacteriocin production through AHL. The arabinose inducible part of the system is present but remains uninduced. b & c) Flow cytometry data showing BFP and GFP expression at varying concentrations of AHL. The grey dashed line shows the median level of fluorescence of the empty *E. coli* MG1655 strain. d, e & f) Time-course data of the SPoCK system at varying AHL induction concentrations.

in GFP expression.

The mild repression of GFP expression shown in Figure 4.11c may indicate that, with the quorum sensing construct on a low copy SC101 plasmid, complete repression is not possible. To determine whether increasing the copy number of the quorum sensing plasmid is able to provide more complete repression, charac-

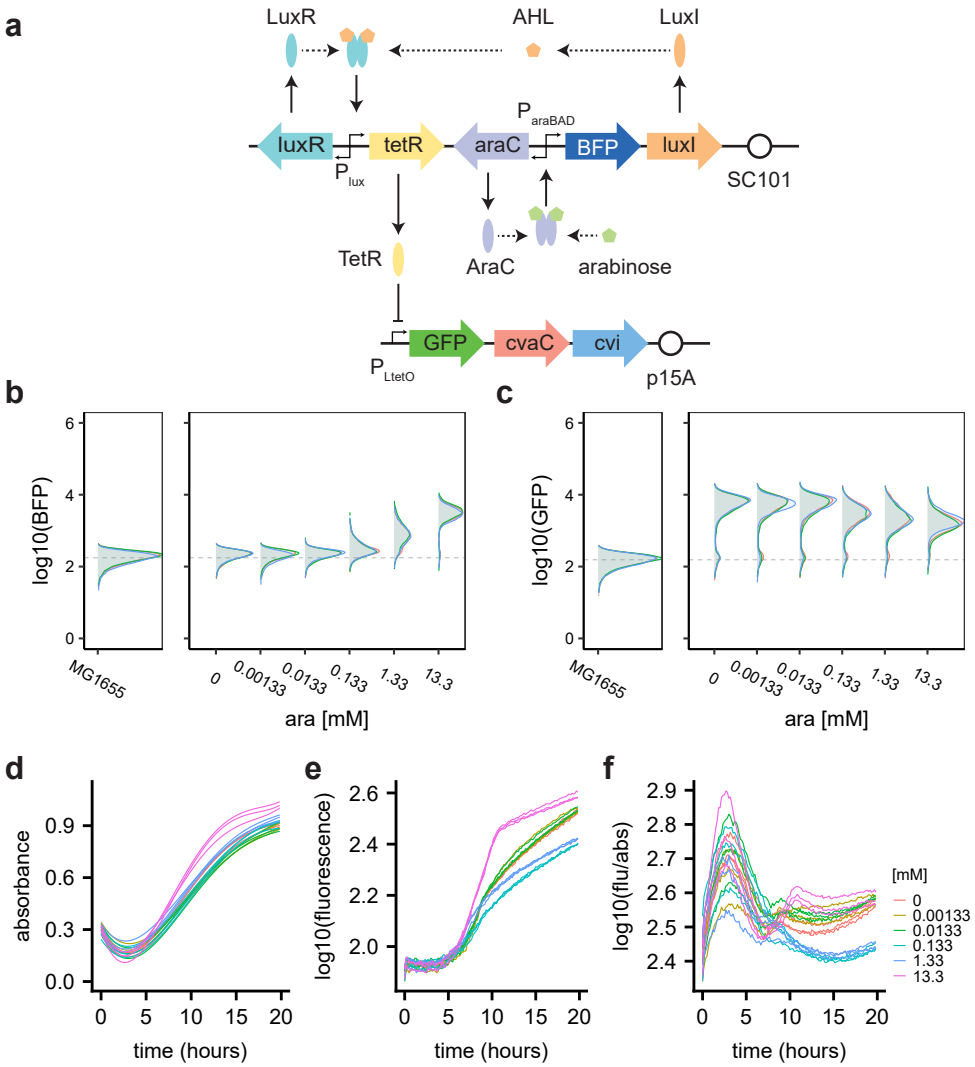

Figure 4.11: Inhibition of bacteriocin production in the SPoCK system with arabinose. a) A cartoon illustrating the mechanism of inhibition of bacteriocin production through arabinose induction of the quorum sensing system. b & c) Flow cytometry data showing BFP and GFP expression at varying concentrations of arabinose. The grey dashed line shows the median level of fluorescence of the empty *E. coli* MG1655 strain. d, e & f) Time-course data of the SPoCK system at varying arabinose induction concentrations.

terisation of the system under arabinose induction was carried out with a variety of origins of replication. The BFP flow cytometry data, Figure 4.12a, shows expression increasing with copy number at all inducer concentrations. Similarly, the GFP data, Figure 4.12b, shows increasing repression with copy number across all inducer concentrations. However, even when uninduced the system shows repression of GFP expression in the higher copy number BR322 and pUC plasmids; indeed

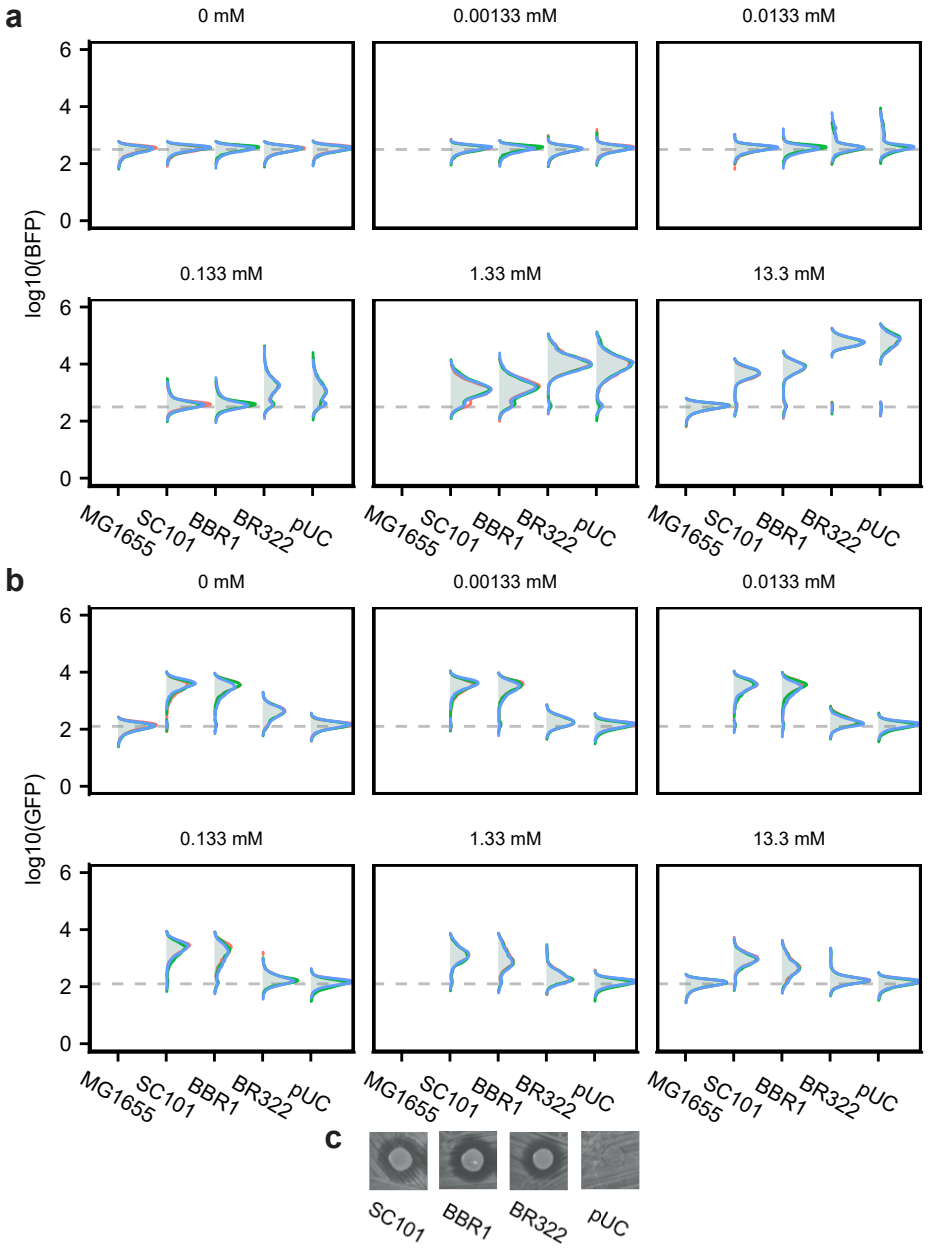

Figure 4.12: SPoCK with different copy number plasmids. a & b) BFP and GFP expression at various arabinose concentrations. The grey dashed line shows the median fluorescence level of empty *E. coli* MG1655. c) Inhibition by uninduced SPoCK strains.

the pUC plasmid shows complete repression. The inhibition zones produced by uninduced SPoCK systems, Figure 4.12c, show that all of the systems can produce bacteriocins except for the highest copy number pUC system, confirming the GFP cytometry data. This also confirms that the bacteriocin production can be turned off, which is crucial for the system to function dynamically.

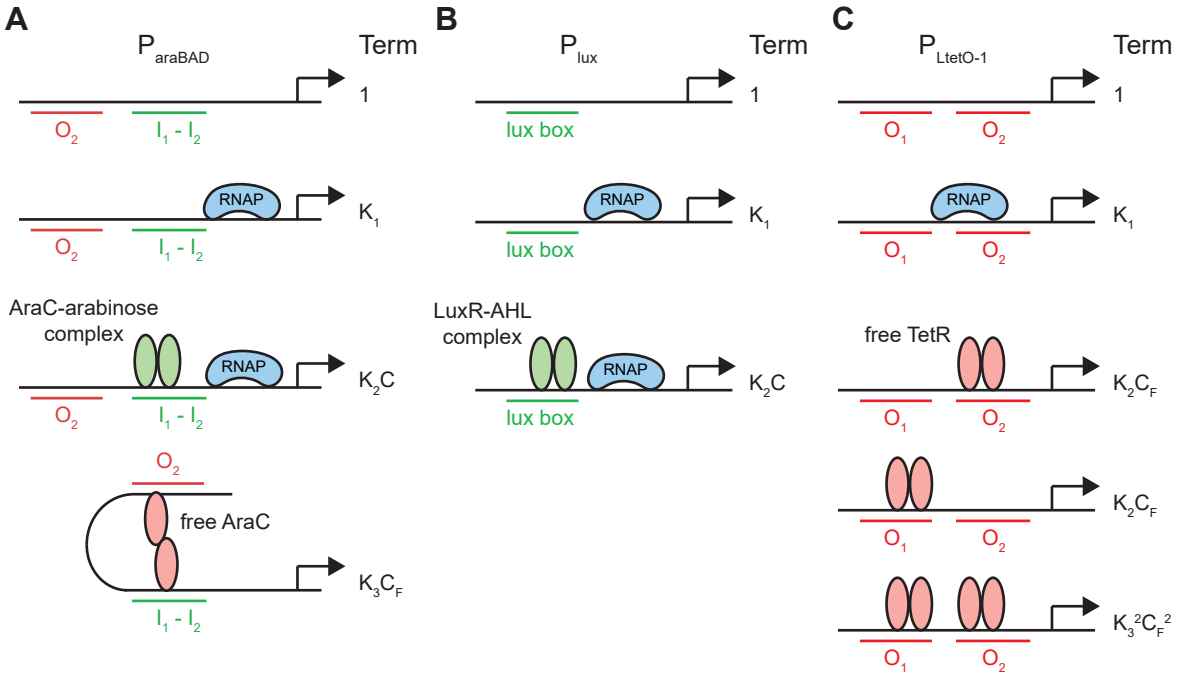

Figure 4.13: States of the SPoCK promoters.

4.4 A More Complete Model of SPoCK Dynamics

The model presented in the previous Chapter contained abstractions of parts of the system that can now be detailed; specifically the production of AHL and bacteriocin by the killer strain. Here, the Shea-Ackers formalism is used to model the expression from promoters in the system. This is chosen over other description of promoter activity such as Hill functions and logistic curves as it offers a more mechanistic understanding of the behaviour of the promoters [162, 105].

4.4.1 Shea-Ackers Characterisation of SPoCK Promoters

The AHL expression rate was previously described by a constant, k_A. In reality AHL is produced by the LuxI protein which is, in this system, expressed from the ParaBAD promoter driven by arabinose. Using the Shea-Ackers formalism for describing promoter activity, a model of the probability of expression from the ParaBAD promoter, with states shown in Figure 4.13A, can be written as:

$$P_B(t) = \frac{K_{1_B} + K_{2_B} C_B}{1 + K_{1_B} + K_{2_B} C_B + K_{3_B} C_{F_B}} \tag{4.1}$$

$$C_B = C_{0_B} \frac{L_B^{n_B}}{K_{d_B}^{n_B} + L_B^{n_B}} \tag{4.2}$$

$$C_{F_B} = C_{0_B} - C_B \tag{4.3}$$

where P_B is the probability of expression from the ParaBAD promoter, C_{0_B}, C_B and C_{F_B} are the total, ligand bound and ligand free concentrations of AraC in each cell, L_B is the concentration of arabinose in the system, K_{d_B} is the dissociation constant of arabinose-AraC, and n_B is the cooperativity of ligand binding. The intra-cellular concentration of the ligands for all transcription factors is assumed to be equal to the environmental concentrations.

In the simple model in the previous Chapter, bacteriocin expression was described by a repressive Hill function with activity dependent on AHL. The system constructed in this Chapter requires an intermediate step to invert the AHL signal; the bacteriocin is repressed not by AHL directly but by AHL activated TetR. The TetR is expressed from the PLux promoter, with states shown in Figure 4.13B, the activity of which can be written as:

$$P_L(t) = \frac{K_{1_L} + K_{2_L} C_L}{1 + K_{1_L} + K_{2_L} C_L} \tag{4.4}$$

$$C_L = C_{0_L} \frac{L_L^{n_L}}{K_{d_L}^{n_L} + L_L^{n_L}} \tag{4.5}$$

$$L_L = A(t) \tag{4.6}$$

where P_L is the probability of expression from the PLux promoter, C_{0_L} and C_L are the total and ligand bound concentrations of LuxR in each cell, L_L is the concentration of AHL in the system, K_{d_L} is the dissociation constant of AHL-LuxR, and n_L is the cooperativity of ligand binding.

Finally, the bacteriocin is expressed from the PLtetO-1 promoter which is re-

pressed by TetR, as shown in Figure 4.13. The expression from this promoter can be written as:

$$P_T(t) = \frac{K_{1_T}}{1 + K_{1_T} + 2K_{2_T}C_{F_T} + K_{2_T}^2 C_{F_T}^2} \tag{4.7}$$

$$C_{F_T} \propto P_L(t) \tag{4.8}$$

where P_T is the probability of expression from the PLtetO-1 promoter and C_{F_T} is the ligand free concentration of TetR in each cell. Since no ligand for TetR is added to the environment, no consideration needs to be made regarding the binding of ligand to transcription factor as has to be done for the other promoters.

Using the flow cytometry data for the activity of each of these promoters after 24 hours of induction, assumed to be at steady state, with a range of inducer concentrations, values for the parameters in the above models can be ascertained. Often these models are fit using regression methods that return point estimates of each parameter [2, 163]. However, by using Bayesian methods it is possible to get a better understanding of the interplay between different parameters in the system through looking at posterior correlations. It also allows one to understand the constraints on particular parameters in order to achieve a given behaviour. Here, I use Stan [164] to produce fits for the flow cytometry data, with a log-likelihood given by:

$$L(\theta|\mathbf{x}) = \sum_{r}^{R}\sum_{i}^{I}\sum_{p}^{P_p} \log(P(x_{r,i,p}|\theta)) \tag{4.9}$$

$$P(x_{r,i,p}|\theta) = N(P_{i,p}, \sigma_p) \tag{4.10}$$

where $i \in I$ are the inducer concentrations, $r \in R$ are the replicates of each measurement, $p \in P_p$ are each of the promoters, θ is the parameter set and $x \in \mathbf{x}$ are the flow cytometry data. The model fits are shown in Figure 4.14 and a correlogram of the parameters is shown in Appendix B. The promoters are fitted concurrently, rather than individually, as the output from the PLux promoter is the input into the

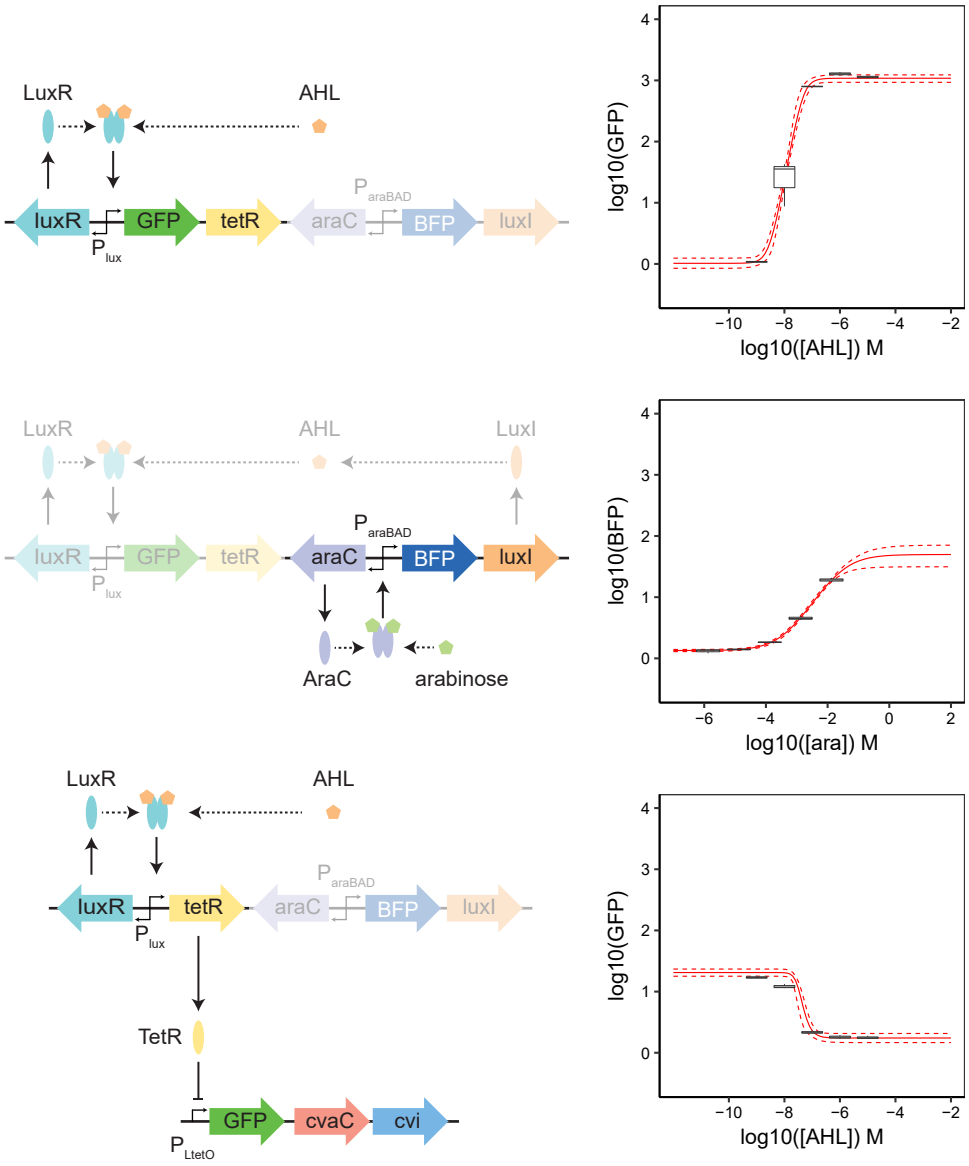

Figure 4.14: Fits of the Shea-Ackers models to flow cytometry data of SPoCK promoter activity.

PLtetO-1 promoter. Unfortunately, concurrently fitting the PLux and PLtetO-1 promoter activities to arabinose induction data was unsuccessful. This is most likely due to this method assuming steady-states have been reached in promoter expression but due to the density dependence of the PLux promoter, steady state may not have been reached.

4.4.2 Finding Equilibria and Oscillations Using Approximate Bayesian Computation

The models of promoter activity can be combined with the simple SPoCK model described in the previous Chapter in order to understand how the system that has been built will behave. The changes to the simple model are small:

$$k_\beta(t) = P_T(t)\, k_{\beta_{max}} \tag{4.11}$$

and

$$k_A(t) = P_B(t)\, k_{A_{max}} \tag{4.12}$$

where $k_{\beta_{max}}$ and $k_{A_{max}}$ are the maximal rates of expression of bacteriocin and AHL. Note that there is a simplification here in that the expression rate of AHL, k_A is modelled as proportional to the activity of the ParaBAD promoter, rather than the concentration of intracellular LuxI. Due to the degradation tag fused to the LuxI, the concentration of LuxI should be closely related to current ParaBAD activity. As such, it is felt that this simplification is a reasonable one.

The effect of the bacteriocin on the rate of killing of competitor cells is also updated. Previously the relationship between bacteriocin concentration and cell death was linear, with a constant ω giving a death rate per unit of bacteriocin. A more realistic approach is used here in which the death rate is related to bacteriocin concentration by a Hill function. This prevents death rates from becoming unrealistically fast if bacteriocin concentration becomes higher than expected. As such:

$$\omega_C(t) = \frac{\omega_{C_{max}} \beta(t)^{n_\omega}}{K_\omega^{n_\omega} + \beta(t)^{n_\omega}} \tag{4.13}$$

Prior distributions for each of the model parameters, Table 4.1, were ascertained either from the literature, from experimental data or from sensible estimates.

Table 4.1: Prior distributions of SPoCK model parameters for ABC simulations

Parameter	Description	Prior	unit
D	Dilution rate	$\mathrm{U}(0.01, 0.5)$	h^{-1}
L_B	Arabinose concentration	$\mathrm{U}(10^{-6}, 10^{-2})$	M
$\mu_{X_{max}}$	Maximum growth rate of killer strain	$\mathrm{U}(0.7, 0.9)$	h^{-1}
$\mu_{C_{max}}$	Maximum growth rate of competitor strain	$\mathrm{U}(0.7, 0.9)$	h^{-1}
$\omega_{C_{max}}$	Maximum death rate induced by bacteriocin	$\mathrm{U}(0.5, 2)$	h^{-1}
K_ω	Half-saturation concentration concerning bacteriocin induced death	$\mathrm{U}(10^{-16}, 10^{-12})$	M
n_ω	Hill coefficient of bacteriocin induced death	$\mathrm{U}(1, 2)$	-
K_X	Concentration of substrate at which killer strain growth is half-maximal	10^{-5}	gL^{-1}
K_C	Concentration of substrate at which competitor strain growth is half-maximal	10^{-5}	gL^{-1}
S_0	Reservoir substrate concentration	4	gL^{-1}
γ_x	Yield coefficient of killer strain	10^{12}	cells g^{-1}
γ_c	Yield coefficient of competitor strain	10^{12}	cells g^{-1}
$k_{A_{max}}$	Maximum expression rate of AHL	10^{-15}	mol h^{-1}
$k_{\beta_{max}}$	Maximum expression rate of bacteriocin	10^{-22}	mol h^{-1}

The priors for the promoter parameters are given by the posterior distributions from the model fitting above.

In order to look for equilibria states, the gradients of the killer and competitor populations at the last time-point were used for the distance function. Examples of equilibria time-series are shown in Figure 4.15 and the posteriors for key parameters are shown in Figure 4.17.

In order to search for oscillatory states a method was used to classify peaks in a simulated time-series by identifying time-points of local maxima. Three tests are used to create a distance function:

1. Are there at least N peaks in the time-series with amplitude above a threshold?

2. Is the amplitude of the final peak above a threshold?

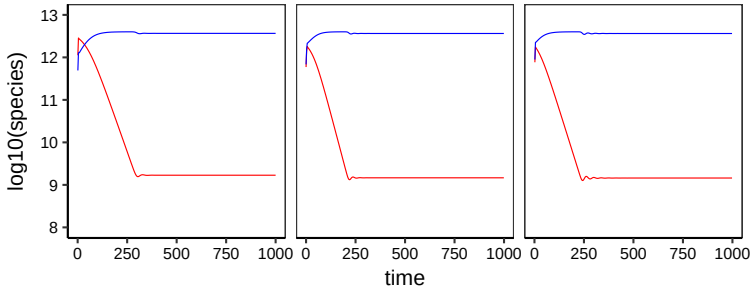

Figure 4.15: Simulated equilibria of the killer (red line) and competitor (blue line) populations in the implemented SPoCK system.

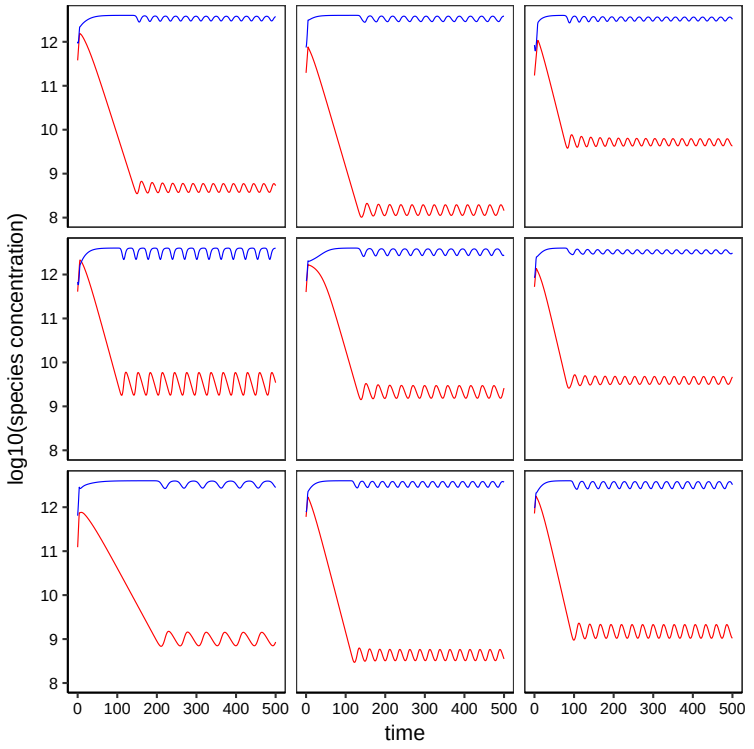

Figure 4.16: Simulated oscillations of the killer (red line) and competitor (blue line) populations in the implemented SPoCK system.

3. Are the amplitudes of the peaks within a threshold standard deviation?

These tests do not necessarily identify stable limit cycle oscillations but do require oscillations to be sustained for a specifiable amount of time. Examples of such time-series are shown in Figure 4.16. The posterior distributions produced for the key parameters by the ABC algorithm are shown in Figure 4.17.

The posterior distributions in Figure 4.17 show clear differences between the stable and oscillatory behaviours. As was seen with the simple model, oscillatory

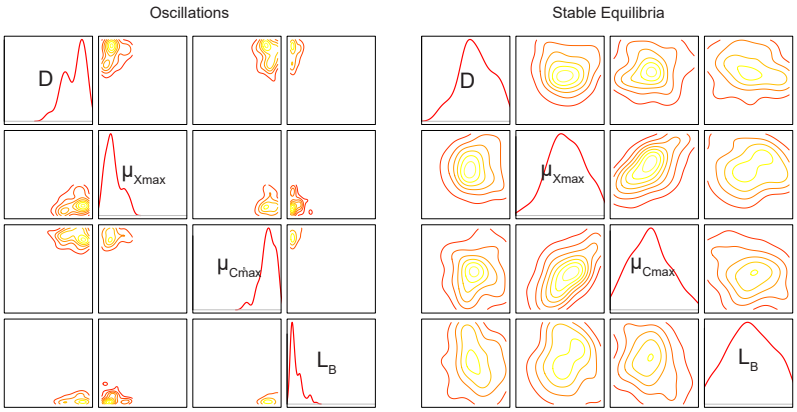

Figure 4.17: Posterior distributions for key parameters to achieve oscillations and equilibria. The posteriors to achieve oscillations are far more constrained than those for stable equilibria.

states require fast dynamics in the system produced by a high dilution rate and high competitor growth rate. Here the dilution rate and competitor growth rate must be high and the arabinose concentration must be at the lower end of the prior range.

4.5 SPoCK in Chemostat

Now that a functioning SPoCK system has been implemented it can be tested in competition. The competitor strain was *E. coli* MG1655 with mCherry expression and gentamicin resistance integrated into the chromosome. An Eppendorf DASbox mini bioreactor was used which allows four experiments to run in parallel with control over dilution rate, gas mixing and pH.

Two chambers were run at a dilution rate 0.1 hr^{-1} and two at 0.05 hr^{-1}. All chambers were inoculated with identical volumes of killer and competitor culture. The single cell fluorescence was measured at several time points across 51 hours using flow cytometry. The data was put through the same analysis pipeline as previously described to select singlet bacteria. Figure 4.18 shows an example of the flow cytometry data for a single chamber at a single timepoint.

The single cells were then clustered on fluorescence, with cells expressing BFP and/or GFP classified as killer and those expressing mCherry or no fluorescence as competitor. The non-fluorescing cells were classified as competitor as it was noted in later experiments that the chromosomally integrated mCherry reporter produced

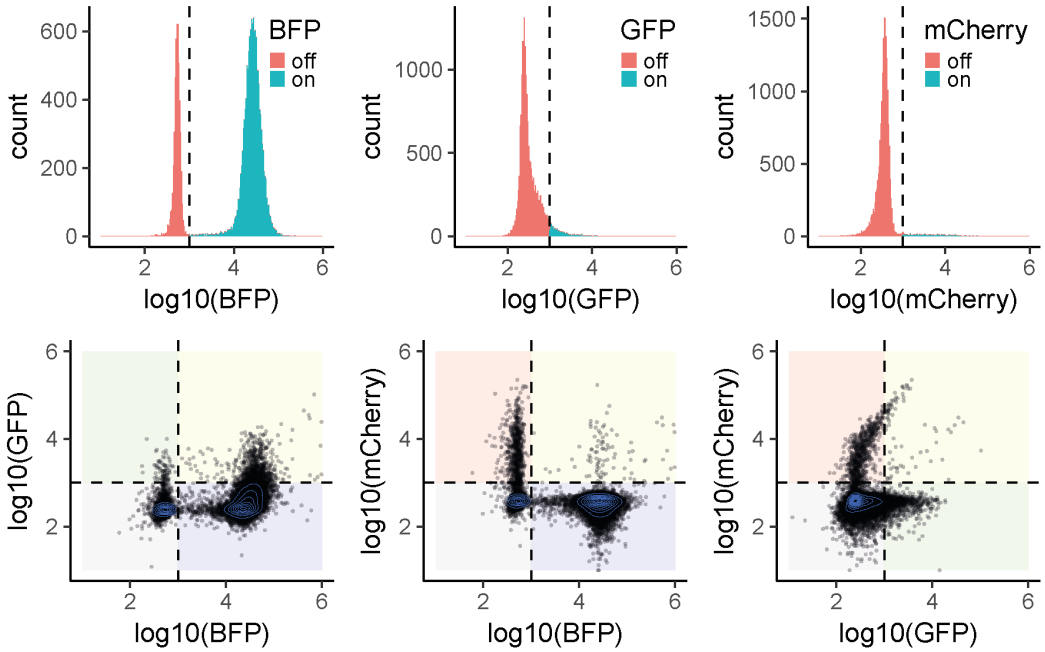

Figure 4.18: Example flow cytometry data for a single chamber at a single timepoint of SPoCK in chemostat. The dashed lines show an arbitrary threshold for classifying cells as on or off for a given fluorescent protein.

very low levels of fluorescence when growing exponentially. The results, in Figure 4.19, show the data from the two chambers diluted at 0.05 hr^{-1}. The killer population initially overtakes the competitor population. After 30 hours, the competitor begins to out-compete the killer and its population increases relative to the killer. Using the parameters inferred previously for the constructed SPoCK system, i.e. those for the Shea-Ackers promoter activities, the model detailed above was fitted to the chemostat data using a using Bayesian methods. Table 4.2 shows the values for the parameter fits.

The model fitting predicts that, under the conditions used here, the killer population will go to extinction after roughly 250 hours. However, using the fitted parameters, a prediction for the outcome of competition between the two strains with no SPoCK system present, shows that the "killer" strain would go to extinction far faster. The model fitting indicates that the SPoCK system causes a very large burden on the killer strain leading to a far lower maximal growth rate, $\mu_{X_{max}} = 0.85$, compared to the the maximal growth rate of the competitor strain, $\mu_{C_{max}} = 2.83$.

Table 4.2: Fitted parameters for SPoCK model to chemostat data. U(x,y) indicates a uniform distribution between x and y.

Parameter	Prior	Mean of Posterior
$\mu_{X_{max}}$	U(0.6, 2.85)	0.85
$\mu_{C_{max}}$	U(0.6, 2.85)	2.83
ω_{max}	$10^{\mathrm{U}(0,9)}$	$10^{0.31}$
K_ω	$10^{\mathrm{U}(-10,-4)}$	$10^{-4.44}$
$k_{A_{max}}$	$10^{\mathrm{U}(-21,-16)}$	$10^{-18.51}$
$k_{\beta_{max}}$	$10^{\mathrm{U}(-23,-17)}$	$10^{-21.58}$
$\beta(0)$	$10^{\mathrm{U}(-15,-3)}$	$10^{-4.37}$
n_ω	U(1, 2.5)	2.46

The two chambers diluted at 0.1 hr^{-1} suffered from pump issues at various points through the experiment but the data can be seen in Appendix C.

The cultures at the end of the chemostat experiment were assayed for the presence of bacteriocin in the media and the ability of the remaining cells to still produce bacteriocin when arabinose is not present. The results can be seen in Figure 4.20. The cultures, from all chambers, that were filtered to remove bacterial cells, produced no inhibition of a susceptible *E. coli* MG1655 strain. This indicates that no bacteriocin is present in the media of any of the chambers at the final time-point. This reinforces the data in Figure 4.19 showing recovery of the competitor strain. The cultures, when unfiltered allow the bacteria present in the culture to grow and, if possible, produce bacteriocin. The cultures in chambers 3 and 4 show zones of inhibition, proving that they still have the ability to produce bacteriocin when not being inhibited. However, the cultures from chambers 1 and 2 do not show any inhibition, indicating that this ability is lost. Further, the morphology of the colony produced from chamber 2 is rather different than that of the other cultures suggesting drastic change in the bacteria present. A similar assay carried out on the cultures from the chambers 24 hours before hand showed that the bacteria in chamber 1 were still capable of producing bacteriocin but those in chamber 2 were already producing a bright colony.

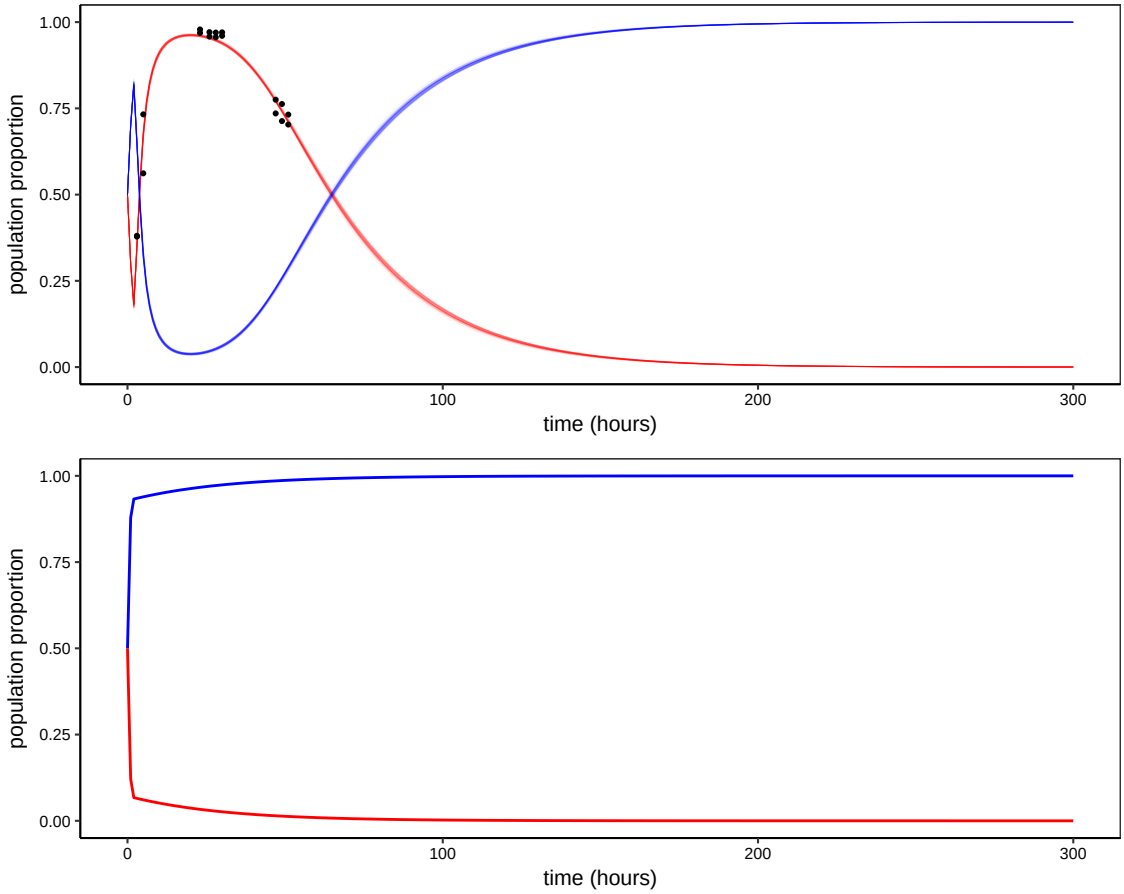

Figure 4.19: SPoCK in a bioreactor with competition between the killer strain (red) and the competitor strain (blue). In the top panel, the black circles show samples from two replicate chambers with dilution rate set to $D = 0.05\text{hr}^{-1}$ that were used to fit the model. The lines show the fitted model. The bottom panel shows a prediction of what would happen to the two strains if the SPoCK system were not present to counteract competitive exclusion.

4.6 Discussion

In this Chapter I have described the construction and characterisation of a system for the density dependant killing of competitors by an engineered bacterial strain. Each element of the system has been characterised using flow cytometry data, Shea-Ackers models of promoter activity and Bayesian model fitting. The system was tested in competition in chemostat and the results, though preliminary, showed behaviour that suggests the SPoCK system is behaving dynamically. Finally, using ABC, conditions were discovered under which one might expect to produce oscillating or stable communities.

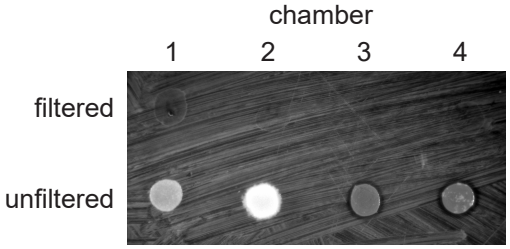

Figure 4.20: Inhibition assay of chemostat cultures at the end of the competition experiment. The top row shows no inhibition when cultures are filtered to remove bacterial cells, indicating no bacteriocin present in the media. The bottom row shows unfiltered cultures. The bacteria growing in chambers 3 and 4 are still able to produce bacteriocin causing inhibition when arabinose is not present. The cultures in chambers 1 and 2 produce no zones of inhibition, suggesting a loss of bacteriocin production capability.

An inducible bacteriocin production system was constructed, the first to my knowledge, using a the TetR repressible PLtetO-1 promoter and the microcin-V bacteriocin. This was shown to function over a range of inducer concentrations; though the GFPmut3b used as an expression level proxy did not provide a useful proxy. However, computational prediction of promoters in the microcin-V cassette found a putative promoter within the *cvi* gene, upstream of the *cvaC* bacteriocin gene. This potentially means that repression of the synthetic promoter cloned upstream of the *cvi* gene may not be enough to repress expression of the bacteriocin. As such, a plasmid was created using synthesised *cvaC* and *cvi* genes, separated and placed in the opposite order to the wild-type operon.

A quorum sensing system was constructed from two parts. The receiver part consisted of a lux promoter with luxR being transcribed in the reverse direction and sfGFP-LAA transcribed in the forward direction. This was demonstrated in a low copy SC101 plasmid to have $\sim$100 fold dynamic range of AHL induction. Time-courses demonstrated that the higher levels of induction had a small but noticeable impact on growth. Further, a transient repression of the promoter was recorded at all inducer concentrations at a similar point in growth. This may be indicative of the presence of one of the known growth phase dependant transcriptional regulators that repress the lux promoter.

The sender part of the the quorum sensing system consisted of an ParaBAD

promoter driving expression of its transcription factor araC in the reverse direction and mTagBFP2 and luxI-LAA in the forward direction. Again, this system was inducible with at least a $\sim$1000 fold dynamic range with arabinose. The growth curves showed that, at the highest concentrations of arabinose, growth rate increased possibly due to utilisation of the arabinose as a carbon source [165]. A plate reader assay could not be used due to a lack of filters to detect BFP but using PBS as a medium, it was shown that the induction of this sender part of the circuit was not density dependant. Unfortunately the same assay couldn't be used to demonstrate the density dependence of the receiver part of the circuit as the lux promoter did not turn on in PBS.

The sender and receiver system were combined on a single plasmid and the ability to induce the receiving lux promoter by arabinose induction of the sending system was demonstrated. The dynamic range on a low copy SC101 plasmid was shown to be $<$ 100 fold with arabinose. BFP proved to be a poor marker for expression from the sender part, with very small increases in expression from the basal level correlated with virtually maximal expression of GFP from the receiver part.

Plasmid copy number was identified as a further manipulable parameter so three additional origins of replication were characterised. The medium copy BBR1 origin of replication behaved in a similar fashion to that of SC101, producing greater fluorescence homogeneity. The high copy BR322 plasmid appeared to be in an off state when uninduced, though the GFP expression distribution showed a tail of GFP expressing cells. However, when induced with arabinose, at a concentration determined from the SC101 system to produce maximal GFP induction, the population split with a large non-functioning population. This was further explored with a range of arabinose concentrations revealing that at low inducer levels the system behaved as desired. A non-functioning sub-population was detectable at the second lowest inducer concentration tested. At the two highest concentrations tested, a third population was detectable in which BFP expression was functioning but GFP expression was not. It is clear that the sender part is malfunctioning in the population without BFP expression. It is less obvious whether, in this population, the

receiver part is also malfunctioning. However, in all of these samples there are populations still producing the fluorescent proteins expected so it can be assumed that there is AHL in the environment. This leads to the conclusion that the receiver part is malfunctioning as well. At the two highest concentrations, there is a third subpopulation in which BFP expression is on but GFP expression is not. This clearly indicates malfunctioning of the receiver only. It is well understood that pushing cells to produce large quantities of proteins exerts a pressure which selects for those cells which manage to accrue loss of function mutations in non-essential operons [32]. The very high, uncontrolled plasmid copy number of the pUC based plasmid produced high levels of GFP and low level BFP expression when uninduced and malfunctioned in a similar way to the BR322 system when induced. The BFP expression when uninduced indicates that the leakiness of the ParaBAD promoter is amplified by the very high copy number. However, some of the BFP expression detected may be spill over from the extremely high GFP fluorescence into the BFP detection channel. As such, it is difficult to determine to what extent the known leakiness of the lux promoter or the supposed leakiness of the ParaBAD promoter contribute to this expression pattern. The experiments with the full SPoCK system using the pUC plasmid indicate that the leakiness of the lux promoter is the dominant factor; though it has been shown earlier that very low levels of BFP expression can result in high GFP expression.

The full SPoCK system, consisting of the quorum sensing and bacteriocin plasmids was tested in *E. coli* MG1655. It was shown that GFP expression, the marker used for bacteriocin production, was repressible by induction of the quorum receiver with exogenous AHL. GFP expression was not completely eliminated, even at the highest AHL concentration, though with no degradation tag on the GFPmut3b its half-life was roughly the same length of time as the induction. Induction of the system with AHL had a small positive impact on growth, indicating that bacteriocin and GFP production are, unsurprisingly, more costly than TetR production.

Arabinose induction of the full SPoCK system resulted in expected BFP expression dynamics. The repression of GFP expression was not as dramatic as when

the system is induced with AHL. However, as noted before, the half-life of GFP-mut3b is long and the additional time required for the population density to increase before repression of the PLtetO-1 promoter occurs would result in a lag before GFP production is turned off.

Plasmid copy number of the quorum sensing plasmid was again explored as a system parameter. The first point to note is that the uninduced pUC plasmid produces no BFP expression, as was seen when characterising the quorum plasmid alone. This lends greater weight to the supposition that the high GFP signal leaked into the BFP channel rather than a great deal of leakiness from the ParaBAD promoter. This is reinforced by the total lack of GFP expression from the bacteriocin plasmid in the uninduced pUC system and lack of inhibition zone on the inhibition assay, indicating total repression of the PLtetO-1 promoter through leaky production of TetR. The differences in expression for each origin of replication correspond to the expected patterns given previous results. Here, though, there is minimal heterogeneity in population behaviour and that which is seen is in the BFP expression from the ParaBAD promoter which is a known binary fate phenomena [166].

The full system was tested in a chemostat with competition. Flow cytometry measurements showed initial removal of the competitor population followed by recovery. Inhibition assays indicated that bacteriocin production can be turned off inside the bioreactor and that the ability to produce bacteriocin is maintained. However, this was not the case in all four experiments. The results should only be viewed as preliminary as there were several weaknesses in the experiment. Firstly, the fluorescence marker used to identify the competitor strain produced weak variably fluorescing cells. As such, non-fluorescing cells were classified as the competitor strain based on an assumption that the killer strain must be fluorescing in either the BFP or GFP channel at all times. Secondly, the DASbox system that was used did not seem to maintain perfect pump accuracy for feed of media and removal of waste. As such, the volume within each chamber varied over the course of the experiment. As such, dilution rates can't be considered to be stable. Finally, the limited number of chambers that can be run in parallel meant that for this preliminary run, no

controls of the competitor and killer alone were made.

Each of the components of the SPoCK system were characterised using flow cytometry data and Bayesian model fitting to Shea-Ackers models. These parameters were then used to make the simple model, shown in the previous Chapter, more representative of the system as implemented. Using ABC, conditions under which one could expect to see oscillating and equilibria states were found. The key parameters to achieve these different states were the dilution rate and growth rate of the competitor and the concentration of arabinose in the system. While the growth rate may not be easily modifiable for some possible applications, the arabinose concentration is an easily changeable parameter and was designed to be used to modify system function. Further, though the dilution rate may be easily changed in laboratory experiments, in an industrial setting dilution rate is likely to have been carefully designed to achieve the goal of the industrial process.

Some weaknesses remain in the model. The production of AHL remains a simplification from the real mechanism which involves the intermediate production of LuxI. Further, no time delay is modelled in the system through the accumulation and degradation of transcription factors and ligands, other than AHL. This means that the instantaneous activity of the PLtetO-1 promoter is dependent on the instantaneous activity of the PLux promoter, not on the accumulated level of TetR in each cell. For the production of AHL this may be a valid abstraction due to the ssrA degradation tag on the LuxI protein. However, the half-life of non-tagged proteins can be considerable and may cause differences to the dynamics modelled here.

4.7 Conclusion

In conclusion, I have produced the first dynamically controllable bacteriocin expression system, with versions that can be induced with the addition of aTc or repressed through the addition of an AHL. Using this system I have developed a tunable density dependent bacteriocin expression system; the first synthetic system of its kind. I have characterised all parts of this system using a Shea-Ackers formalism and a novel Bayesian fitting approach. Finally, I have made the first steps towards demon-

strating the abilities of this system to dynamically control a two species community in a chemostat.

4.8 Future Work

The SPoCK system, detailed in Chapter 3 and constructed in this Chapter, was shown theoretically to produce stable coexistence of two bacterial populations. The system was tested in a chemostat and the results demonstrated the ability for the killer strain to compete against the competitor strain and for the competitor to rebound once bacteriocin production was turned off in the killer. However, the limited number of chambers in the chemostat and reliability problems meant that the results should only be viewed as preliminary. I have begun the construction of an 8 chamber chemostat/turbidostat [167]. This will allow for the concurrent testing of the system along with suitable controls and a reasonable number of replicates; something that was not possible with the equipment available. An alternative approach would be to use microfluidics, as was done with another system for producing stabilised populations [102].

In addition to further experimentation with the currently constructed system, adaptations to that system can be made. One adaptation was the addition of self-regulation, as modelled at the end of Chapter 3. This can be done by adding a toxin under the control of the Plux promoter. It is likely that the translation rate of the toxin gene would need to be carefully tuned in order to prevent unwanted cell death.

Chapter 5

Stabilisation of Synthetic Circuits with Post-Segregational Killing

Manipulation techniques for creating DNA based synthetic circuits have allowed the creation of a wide variety of functional synthetic biological system with applications from industrial scale chemical production through to live bacterial therapeutics. The DNA that encodes these circuits is often developed using plasmids. This is usually due to the fact that plasmids have been widely used historically in molecular and cellular biology research, which has lead to a plethora of methods for their isolation and manipulation [23, 24]. Further, plasmids offer the modularity that has been one of the main aims of the field of synthetic biology since its inception [25, 3]. Finally, it has been shown that circuit copy number is an important parameter in the functioning of genetic circuits [22]. Transferring a system developed on a multi-copy plasmid to a single location within the chromosome while maintaining identical function is not trivial [22]. Inserting genetic circuits into the chromosome of bacterial species is not necessarily simple, though new methods using lambda-Red [16] and CRISPR-Cas9 [17] are making the process easier. The location in the chromosome one inserts genetic circuits can also have a big impact on their function and in bacteria is likely to disrupt native cell function to some degree [20, 21].

However, ensuring the stable inheritance of plasmids when cells divide is a fundamental problem when using plasmids. Plasmid free cells can be produced if plasmids are distributed unevenly within a cell at division. Since the majority of

synthetic circuits carried on plasmids produce a burden on their host cells, the new plasmid-free cell will outgrow the plasmid-bearing cells in the environment. This can quickly lead to the effective removal of the plasmid-bearing population, and the functioning circuit, from the environment [29]. For synthetic bacterial systems to be of any use, one must be able to ensure the predictable presence of that system within its intended environment. It should be noted here that there is a second type of plasmid stability; structural stability. This refers to the faithfulness with which the DNA sequence of the plasmid remains unaltered across rounds of replication. This form of stability is not the topic of this thesis. However, it is of fundamental importance to the continued function of synthetic genetic circuits which is a concern of this work.

Antibiotic selection of bacteria containing plasmids carrying the corresponding antibiotic resistance genes is commonly used in a research environment. However, antibiotics are not used in industrial fermentation due to the financial impact of removal and deactivation [33]. Antibiotics are also unsuitable for clinical applications for a number of reasons including horizontal gene transfer of resistance genes [34, 35] and the disruption of the native microbiota [36]. In light of these limitations, efforts have already been made to re-use a variety of existing microbial machinery to ensure plasmid persistence in more complex environments [33, 79]. Successful alternatives have been demonstrated with the use of toxin-antitoxin (TA) systems [37, 13], active-partitioning mechanisms [13, 38] and auxotrophy [40, 39].

In this Chapter I will describe the use of two toxin-antitoxin (TA) systems and one bacteriocin system to stabilise burdensome plasmids in a therapeutically relevant strain of *E. coli*, both *in vitro* and in an *in vivo* mouse tumour model. As described in Chapter 1, TA systems are a type of post-segregational killing mechanism that work through the production of a plasmid encoded long lasting toxin and its shorter lived antitoxin [76, 77]. In this Chapter I will focus on two TA systems; the commonly used *E. coli* native hok/sok and the axe/txe system isolated from a gram-positive bacteria, *E. faecium*.

Hok/sok is a type I TA system originating from the *parB* locus of the *E. coli*

plasmid R1 that has been shown to be effective in its native *E. coli* as well as the gram-negative *Pseudomonas putida* [124]. It consists of three genes: the host-killing (hok) gene produces the toxic protein which causes cell membrane permeabilisation [168], the suppressor-of-killing (sok) gene produces an RNA which prevents the translation of the hok mRNA, and the modulation-of-killing (mok) gene which is required for hok translation. Figure 5.1a shows the layout of the sequences within the system. A forward promoter drives the transcription of the mok-hok mRNA and a reverse promoter drives the transcription of the sok RNA. In the full length mok-hok mRNA the 5' and 3' ends are the inverse of each other leading to a blunt-end pairing and a stably folded structure with a half-life of 20 minutes [169]. In this folded form the ribosome binding sites are hidden so the mRNA is inactive. Only after processing of the 3' end with RNase II to remove a short nucleotide sequence, and following refolding, does the mRNA become active. Although the translated mok protein does not seem to have any known function, the translation of mok and hok are coupled in such a way that without translation of mok, there is no translation of hok [170]. The sok RNA, with a very short half-life of 30 seconds, is antisense to a region on the mok-hok mRNA which includes the ribosome binding site for mok. This region only becomes exposed once the mok-hok mRNA becomes active. The binding of the sok RNA to the mok-hok mRNA prevents the transcription of mok and, indirectly, hok. The RNA duplex is quickly cleaved by RNase III [171]. This process is shown in Figure 5.1b.

Axe/txe is a proteic, type II TA system originating from the *axe-txe* locus of the gram-positive *Enterococcus faecium* plasmid pRUM [125]. The txe toxin is an endoribonuclease which cleaves mRNA, preventing cells from producing proteins [173]. It has been used to stabilise luminescent reporter plasmids in gram-positive *E. faecium* [125], *E. faecalis* [174], and *B. thuringiensis* as well as in gram-negative *E. coli* [125]. The axe/txe system had been found in 75% of isolates of the common hospital pathogen vancomycin-resistant enterococci (VRE) [175].

A potential alternative to TA systems are bacteriocins, which are bacterially secreted peptides with a bactericidal affect on either a narrow or broad spectrum of

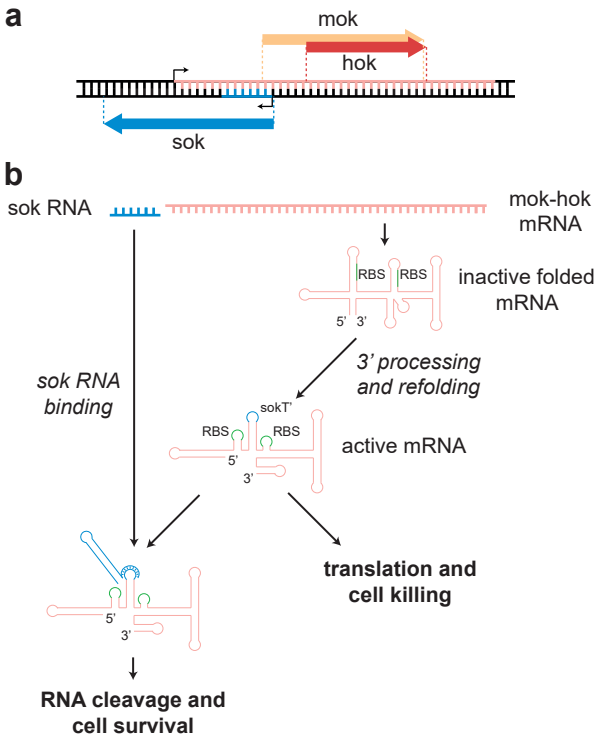

Figure 5.1: The hok/sok toxin-antitoxin system. a) The layout of the hok, sok and mok sequences. Forward and reverse transcription start sites are indicated with black arrows, transcribed sequences are shown as coloured sequences and the putative genes are shown with arrows above and below the double stranded DNA. b) Functional process leading to cell killing or survival. Adapted from [172].

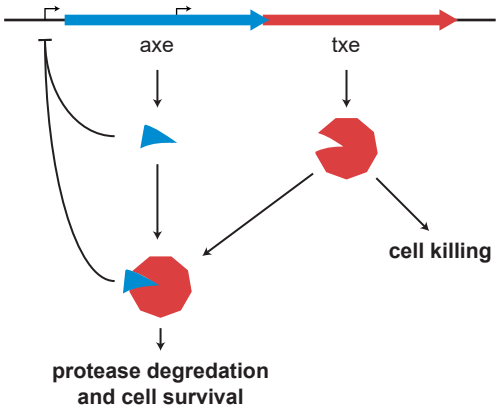

Figure 5.2: The axe/txe toxin-antitoxin system. The layout of the axe and txe genes is shown with the putative transcriptional start sites indicated with black arrows. The axe antitoxin and the axe-txe complex act to repress the promoter upstream of the axe gene.

other bacteria lacking immunity [176]. The secretion of these antimicrobial proteins in to the environment acts to police the plasmid-bearing population, prevent-

ing plasmid-free cells which lack immunity from growing. Bacteriocins, such as the Lcn972 system in *L. lactis* [89] and the colicins A and E2 in *E. coli* [177], have already been used to stabilise plasmids. A potential benefit of bacteriocins over TA systems is their increased robustness to mutational instability. A mutation in the production of a toxin in a TA system would remove the selective pressure on that cell, leading to the loss of the plasmid. However, the selective pressure in a bacteriocin producing population is maintained by the population, even when single cells stop producing the toxin. The only requirement is that some cells maintain bacteriocin production and all cells encode immunity to avoid death [177]. It has already been demonstrated that *E. coli* Nissle carries two bacteriocins, microcins H47 and M, in its chromosome [178]. Additionally, these bacteriocins have been determined to be vital in mediating inter- and intra-species competition in the inflamed gut [179].

In this work I use the *E. coli* native microcin-V system, which has not been characterised as a plasmid-stability mechanism before. The toxin is a low molecular weight, pore forming peptide which is secreted from the producing cell by an ABC-transporter [180]. There is only one small post-translational modification, a disulfide bond in the C-terminal region [180], but it does not require any extra enzyme encoding genes. The dynamics of production from the putative promoters has only been partially explored and it is understood that *fur* binding sites within the promoters lead to repression when iron is plentiful in the environment [181]. The bacteriocin gene itself has a modular structure which has allowed its hybridisation with other bacteriocins, enabling targeting of other bacterial strains [142, 143]. The transport mechanism is an ABC-transporter with two of the genes encoded within the microcin-V cassette and the third, TolC, endogenous to *E. coli* [182, 183]. A third transport protein, cvaA*, maybe transcribed from within the cvaA gene, which stabilises the secretion of the bacteriocin [184, 183]. The transport mechanism of the *E. coli* Nissle chromosomal microcin H47 is similar enough to that of microcin-V that it can secrete the active microcin-V bacteriocin [180], suggesting that the large transport genes do not need to be encoded on the plasmids. The bacteriocin

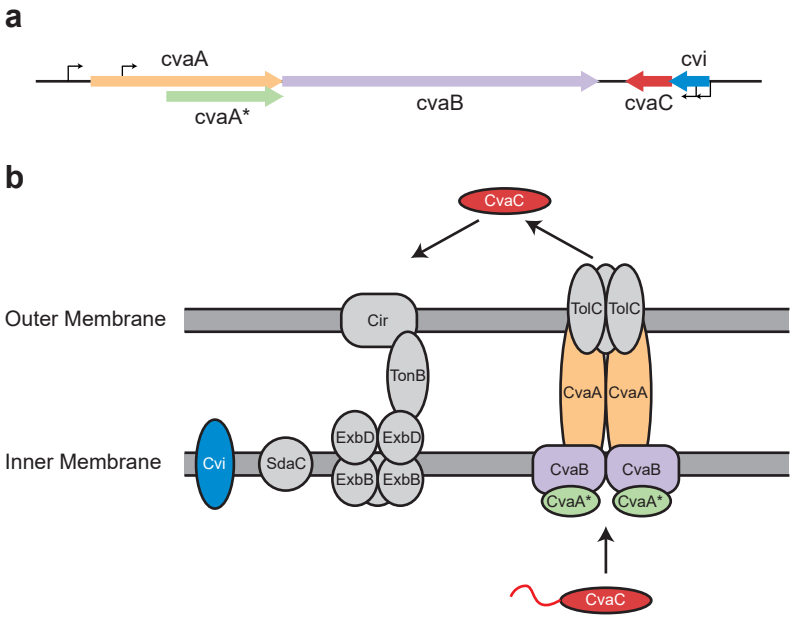

Figure 5.3: The microcin-V bacteriocin system. a) The layout of the cvaA and cvaB transport genes as well as the putative shortened cvaA* gene. The cvaC toxin and cvi immunity genes are transcribed in the reverse direction. Putative transcriptional start sites are indicated with black arrows including one within the cvi gene that has not previously been annotated. b) The transport of the bacteriocin out of the cell through the ABC-transporter and entry in to the cell using the outer membrane receptor Cir and TonB.

recognises the Cir outer membrane receptor [182] and may also use it for translocation in to the cell [185]. It is thought that the TonB-ExbB-ExbD complex, used for iron uptake along with Cir, is necessary for transporting CvaC across the periplasm to allow pore formation in the inner membrane [186]. It has also been shown that a protein involved in serine uptake, sdaC, is important for the insertion of CvaC into the inner membrane [187].

In this Chapter I develop a mathematical model of plasmid loss and use simulations to show how the toxin-antitoxin and bacteriocin systems enhance plasmid stability. Three plasmid stability mechanisms are tested, using the probiotic host strain *E. coli* Nissle, *in vitro* and in an *in vivo* mouse tumour model to determine their stabilising ability in real settings. Bayesian methods are used to infer the killing efficacy and metabolic burden of the three mechanisms.

5.1 A Model of Plasmid Stability with Post-Segregational Killing

An early model of plasmid stability included two populations, plasmid-bearing and plasmid-free, growing exponentially at different rates with a constant probability of plasmid loss from the plasmid-bearing population [188]. The model can be written as a set of ODEs:

$$\frac{dX^+}{dt} = k^+X^+ - \lambda k^+X^+$$
$$\frac{dX^-}{dt} = k^+\lambda X^+ + k^-X^-$$

where, (X^-) and (X^+) are the number of plasmid-free and plasmid-bearing cells, k^+ and k^- are the growth rates, and λ is the loss frequency per cell per generation.

This model was further explored with a particular interest in applying it to the determination of plasmid loss rates from experimental data [189, 190]. The model was also extended to describe plasmid-loss within a chemostat in which growth rates depend on a substrate [191]. A more detailed, biologically accurate model was devised that takes into account the age distribution of bacteria within a population, although the predictions produced were virtually indistinguishable from simpler models [192]. However, this model did highlight the difficulties of trying to infer the plasmid loss rate from measured population data as, due to the exceedingly small plasmid loss rates of most systems, the growth rate differences dominate the dynamics and lead to over-estimation of plasmid loss [192].

A simple model of plasmid loss, assuming a random spatial distribution of n plasmids within a cell, describes plasmid segregation within a dividing cell as a series of n Bernoulli trials [29]. This leads to a probability of producing a plasmid-free daughter cell at division as $p_0 = 2^{(1-n)}$. The assumption of random spatial distribution is important in this model and it was shown that the formation of plasmid dimers changes the plasmid-loss probability [29, 69]. Indeed, dimer formation can prove so bad for plasmid stability that mechanisms for multimer resolution have evolved to prevent the "dimer catastrophe" [69, 73]. The plasmid-loss probability

model would therefore suggest that greatly increasing plasmid copy number would lead to ever greater plasmid stability. However, models have been used to show how plasmids carry mechanisms to prevent excessive plasmid replication in order to maintain stability without too great a metabolic effect on the host bacteria [68].

Here, I extend the early plasmid loss model [188] to include post-segregational killing by a toxin-antitoxin system. In this model there are two populations; plasmid-bearing (X^+) and plasmid-free (X^-):

$$\begin{aligned}\frac{dX^+}{d\tau} &= \underbrace{\gamma X^+}_{\text{growth}} - \underbrace{\lambda\gamma X^+}_{\text{plasmid-loss}} \\ \frac{dX^-}{d\tau} &= \underbrace{X^-}_{\text{growth}} + \underbrace{\lambda\gamma X^+}_{\substack{\text{new plasmid}\\ \text{free cells}}} - \underbrace{\omega\lambda\gamma X^+}_{\text{PSK}}\end{aligned} \tag{5.1}$$

where λ is the probability of producing a plasmid-free daughter cell when a plasmid-bearing cell divides, γ is the ratio of plasmid-free generation time to plasmid-bearing generation time, and ω is the probability of successful post-segregational killing. The time-step τ is equal to one plasmid-free generation. This model shows that, with every plasmid-free generation the extant plasmid-free cells double and new plasmid-free cells are produced by plasmid-loss from the plasmid-bearing population but some of those newly created plasmid-free cells are killed through PSK.

An analytical equation for the proportion of plasmid-bearing cells, ϕ, at time-point t, is given by:

$$\begin{aligned}\phi_t(\lambda,\gamma,\omega) &= \frac{X^+(t)}{X^+(t)+X^-(t)} \\ &= \frac{e^{(\gamma-\gamma\lambda)\tau}}{\nu e^{\tau} + e^{(\gamma-\gamma\lambda)\tau} + \frac{(e^{\tau}-e^{(\gamma-\gamma\lambda)\tau})\gamma\lambda(1-\omega)}{1-\gamma+\gamma\lambda}}\end{aligned} \tag{5.2}$$

where $\nu = \frac{X^-(0)}{X^+(0)}$ and τ is the number of plasmid free generations at time t. For

the purposes of the plasmid loss curves it is assumed that the initial population is entirely plasmid-bearing, $\nu = 0$. Further, in the plasmid loss experiments, detailed later, the populations are grown for 24 hours after which they are diluted 500 fold. As such the simplifying assumption is made that there are ln(500) plasmid-free generations per passage so $\tau = \ln(500)\frac{t}{24}$.

Bacteriocins do not produce post-segregational killing in the traditional sense. Instead there is a constant killing pressure on the plasmid-free population rather than specific killing of newly plasmid-free cells. As such, the equation for the change in plasmid-free population differs to take in to account the different point at which the bacteriocin causes the death of plasmid-free cells:

$$\frac{dX^-}{d\tau} = X^- + \lambda\gamma X^+ - \underbrace{2\omega X^-}_{\substack{\text{bacteriocin} \\ \text{killing}}} \tag{5.3}$$

The analytical solution for the bacteriocin model is:

$$\phi_t(\lambda,\gamma,\omega) = \frac{e^{(\gamma-\gamma\lambda)\tau}}{\nu e^{(1-2\omega)\tau} + e^{(\gamma-\gamma\lambda)\tau} + \frac{(e^{(1-2\omega)\tau} - e^{(\gamma-\gamma\lambda)\tau})\gamma\lambda}{1-\gamma+\gamma\lambda-2\omega}} \tag{5.4}$$

The effects of varying each of these model parameters can be seen in Figure 5.4. The plasmid loss parameter, λ, has an effect on the gradient early on. However, once the first plasmid-free cells arise it is the growth parameter, γ, that has most effect. This can be seen in the strong effect that small changes in γ has on the shape of the curves. When $\gamma = 1$, the plasmid-bearing and plasmid-free cells are growing at the same rate and the only change in population proportion comes plasmid loss, which can be seen in a very shallow gradient. The death parameter, ω, in the TA model has a similar effect to the plasmid loss parameter, λ. It has very little effect once plasmid-free cells have escaped; its effect is lengthening the time before those cells do arise. Intuitively, this is a similar effect to reducing the loss probability, λ. The death parameter in the bacteriocin model has a more drastic effect than in the TA model and doesn't require order of magnitude variation for those changes to be visible. Over a single order of magnitude for the bacteriocin killing the plasmid-loss

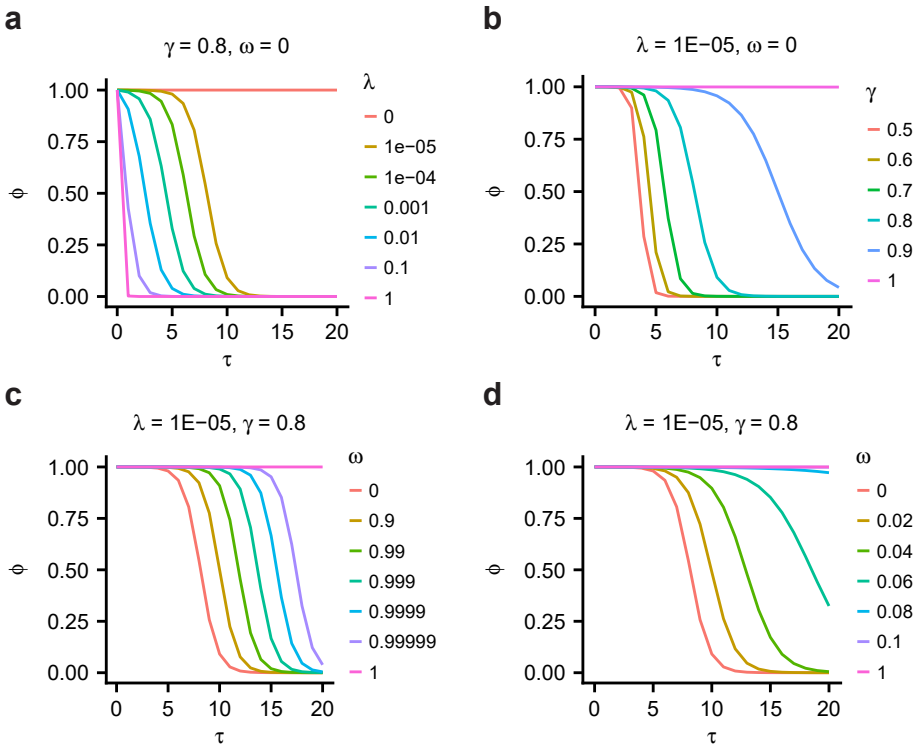

Figure 5.4: Effects of model parameters on plasmid-loss curves. a) Varying the plasmid-loss probability, λ, and b) the doubling time ratio, γ, without post-segregational killing or bacteriocin killing. These are identical for the TA and the bacteriocin models when $\omega = 0$. c) Varying the efficacy of the toxin, ω, in the TA model. d) Varying the efficacy of the bacteriocin, ω, in the bacteriocin model.

curve changes from virtually total plasmid maintenance to almost no extra plasmid stabilising effect.

5.1.1 Modelling Noise in Plasmid Loss Curves

The experimental method used to determine plasmid stability, the protocol for which is detailed in Chapter 2, relies on sampling a population of cells and classifying each one as plasmid-bearing or plasmid-free. Every one of the n cells sampled is equivalent to a Bernoulli trial where "success" is a cell being plasmid-bearing and the probability parameter corresponds to the proportion of plasmid-bearing cells in the culture from which the sample was drawn. As such, a binomial distribution can be used to describe the number of plasmid-bearing cells in any sample:

$$X^{+} \sim \text{Binomial}(n, \phi) \tag{5.5}$$

where n is the number of cells sampled and ϕ is the true proportion of plasmid-bearing cells in the population from which the sample was drawn. In fact, using flow cytometry, the sampling is performed without replacement and as such a hypergeometric model is a more apt description. However, since the fraction of the total population being sampled is very small, $\sim 1/5000$, a binomial model is a reasonable approximation.

The binomial model can describe the noise introduced in sampling the plasmid loss experiments but does not incorporate any intrinsic biological noise, which will result in overdispersion of the data. The "true proportion of plasmid-bearing cells", ϕ, from which the sample was taken, is the parameter that is affected by this biological noise. The uncertainty in this parameter is traditionally modelled using a beta distribution and the binomial and beta model can be combined into a beta-binomial model:

$$X^+ \sim \text{Beta-Binomial}\,(n, \alpha, \beta) \tag{5.6}$$

where

$$\alpha = \phi\left(\frac{1}{\rho} - 1\right) \tag{5.7}$$

$$\beta = \frac{\alpha - \alpha\,\phi}{\phi} \tag{5.8}$$

thus the parameter ρ is used to reflect the uncertainty in ϕ. Simulated plasmid loss curves can be produced by sampling r replicates from the beta-binomial model at t timepoints. The plasmid-bearing population fraction is then given by:

$$\hat{\phi}_{r,t}(\lambda, \gamma, \omega, \rho, n) = \frac{X^+_{r,t}}{n} \tag{5.9}$$

where $X^+_{r,t}$ are r replicate samples at t timepoints from the beta-binomial distribution in Equation 5.6. ϕ in Equations 5.7 and 5.8, is generated at each timepoint using either the TA model in Equation 5.2 or the bacteriocin model in Equation 5.4. The effects of the sample size, n, and the uncertainty parameter, ρ, are shown in Figure

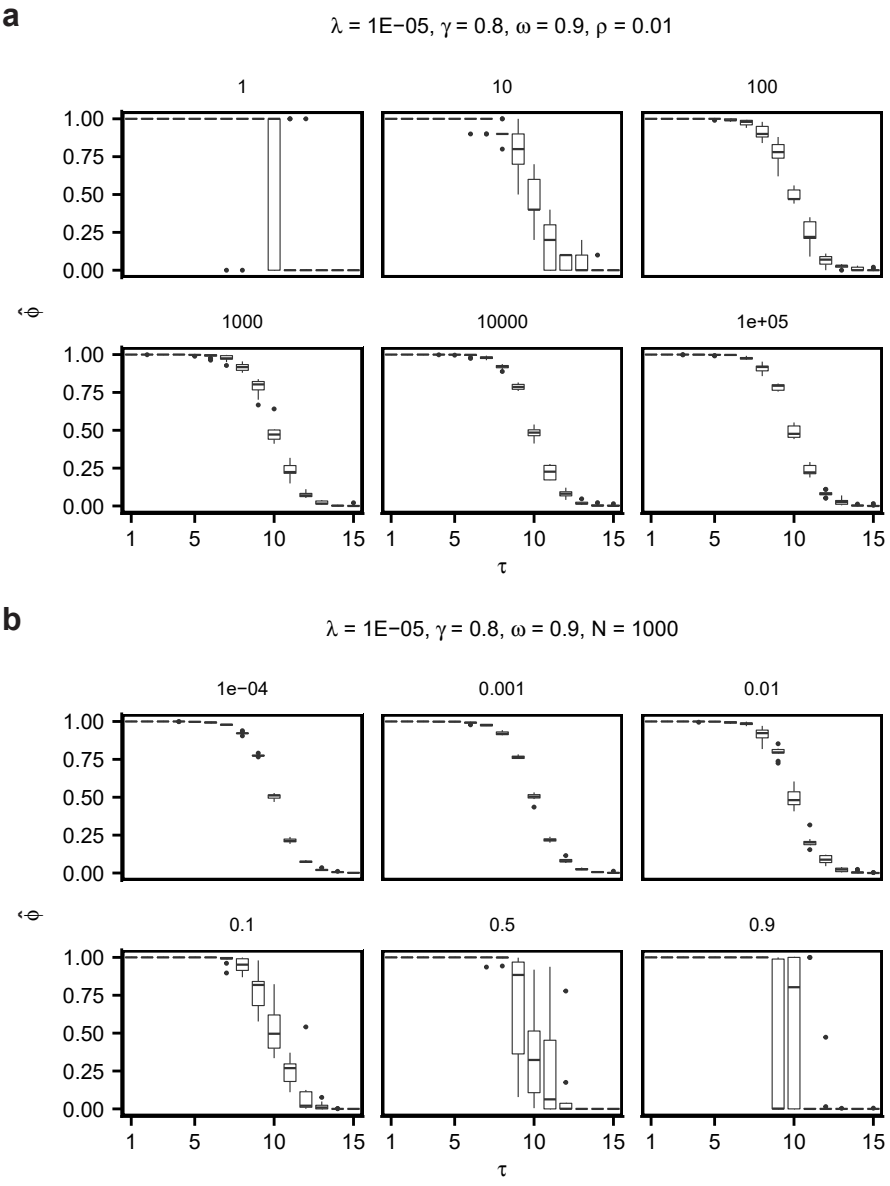

Figure 5.5: Effects of noise parameters on plasmid-loss curve simulated with the TA model. Nine replicates for each plasmid-loss curve. Varying a) sample size, N, and b) the uncertainty parameter, ρ.

5.5.

5.1.2 Bayesian Model Fitting to Plasmid Loss Curves

The models detailed by Equations 5.2 and 5.4 can be fit to experimental plasmid loss curves in order to understand the differences between the tested plasmid stability mechanisms. Here I use a Bayesian method of parameter inference to produce posterior distributions for each parameter, rather than point estimates.

At each timepoint, t, there are r replicates for which analysis of the flow cytometry data produces the number of plasmid-bearing cells $X^+_{r,t}$ in a sample of size $n_{r,t}$. Note that the sample size varies for each sample as all cells within a specified volume are measured rather than cutting off after a given number of events have been recorded.

The aim here is to determine the set of model parameters, $\{\lambda, \gamma, \omega\}$, and noise parameter, ρ, which are most likely to produce the data observed using flow cytometry. Using the beta-binomial model, outlined previously, the log likelihood for each strain can be written as:

$$L\left(\lambda, \gamma, \omega, \rho | \mathbf{X}^+\right) = \sum^{r}\sum^{t} \log\left(P\left(X^+_{r,t} | \lambda, \gamma, \omega, \rho\right)\right) \tag{5.10}$$

where

$$\begin{aligned} P(X^+_{r,t} | \lambda, \gamma, \omega, \rho) &= \text{Beta-Binomial}\left(n_{r,t}, \alpha_t, \beta_t\right) \\ \alpha_t &= \phi_t\left(\lambda, \gamma, \omega\right)\left(\frac{1}{\rho} - 1\right) \\ \beta_t &= \frac{\alpha_t - \alpha_t\ \phi_t\left(\lambda, \gamma, \omega\right)}{\phi_t\left(\lambda, \gamma, \omega\right)} \end{aligned}$$

Markov chain Monte Carlo (MCMC) algorithms and the Stan statistical modelling package [164] are used to produce posterior distributions for each of the model parameters, λ, γ and ω, and the fitting parameter ρ. Two chains of 2000 iterations were produced, with the first 1000 being discarded as burn-in. Convergence was verified by observation that all $\hat{R}$ values were close to 1.

Simulated data is used to determine the effectiveness of this method for inferring true parameter values. Figure 5.6 shows that, for simulated data without post-segregational killing i.e. $\omega = 0$, this method is able to capture the true parameters even with a large amount of noise. This is true even though very broad uniform priors were used; $\log10(\lambda) = U(-10, 0), \gamma = U(0, 1), \rho = U(0, 1)$.

When fitting to the flow cytometry data, it is possible to produce a prior es-

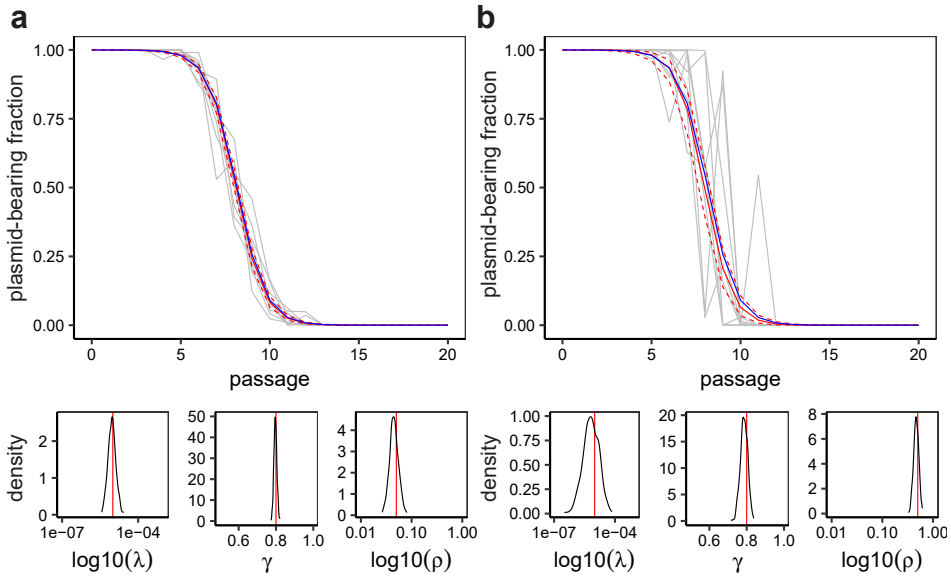

Figure 5.6: Bayesian fitting of simulated data with noise parameter ρ equal to a) 0.05, b) 0.5. Grey lines show each of the 9 simulated trajectories, the blue line shows the true curve, the solid red line shows the median model fit and the dashed red lines show the 95% confidence intervals. Simulated data produced with $\lambda = 10^{-5}, \gamma = 0.8, \omega = 0, N = 1000$. Priors for fitting were $\log10(\lambda) = U(-10,0), \gamma = U(0,1), \rho = U(0,1)$.

timate for the γ parameter, the ratio of plasmid-bearing to plasmid-free generation time, by measuring growth rates of plasmid-bearing and plasmid-free cells.

$$p(\gamma|D_g) = \frac{p(\mu^+|D_g^+)}{p(\mu^-|D_g^-)}$$

By inspection of the posterior distributions, Figure 5.7, one can see that using a normal prior for γ when fitting to simulated data yields little benefit. This suggests that this methodology may be used to ascertain growth rate ratios from competition curves as an additional check to, or instead of, growth curve data.

When fitting the model to the flow cytometry data from the control plasmid, in which there is no PSK mechanism, the PSK death probability parameter ω can be set to 0. This enables the removal of ω from the model fitting which reduces the effects of any potential parameter correlation between ω and the other parameters which may lead to unidentifiability of the model. An assumption can be made that carrying a PSK mechanism does not significantly reduce plasmid copy number [188] which means that the plasmid loss probability is unlikely to change. This

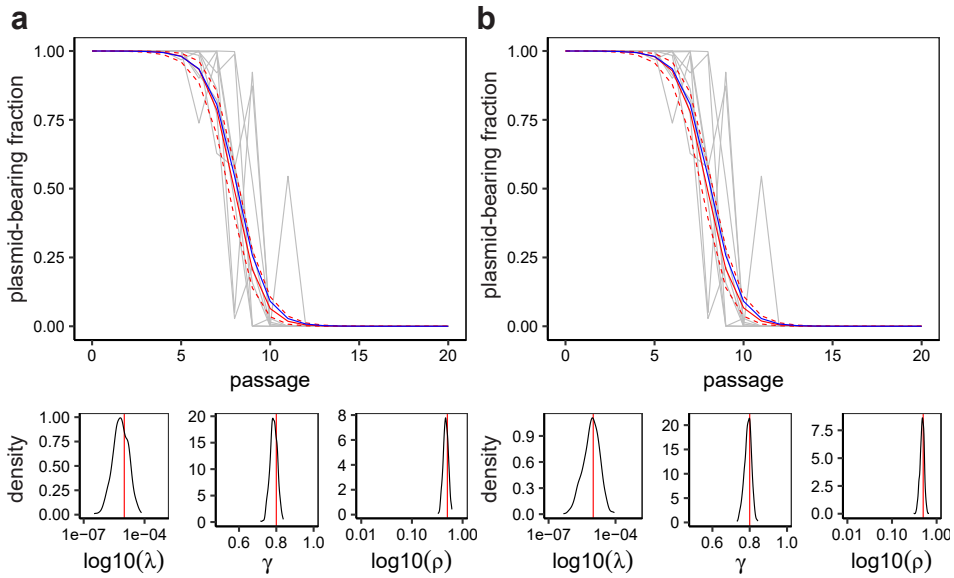

Figure 5.7: Bayesian fitting of simulated data with a) a broad uniform prior $\gamma = U(0,1)$ or b) a tighter normal prior $\gamma = N(0.8, 0.05)$. Grey lines show each of the 9 simulated trajectories, the blue line shows the true curve, the solid red line shows the median model fit and the dashed red lines show the 95% confidence intervals. Simulated data produced with $\lambda = 10^{-5}, \gamma = 0.8, \omega = 0, \rho = 0.5, N = 1000$. Other priors for fitting were $\log10(\lambda) = U(-10,0), \rho = U(0,1)$.

allows for the posterior λ, ascertained from fitting the control plasmid with $\omega = 0$, to be used as a prior for λ when fitting the plasmids carrying PSK mechanisms. The importance of producing a prior estimate for λ can be seen in Figure 5.8. Without a tight prior, correlation between the λ and ω parameters leads to very poor inference of the true parameters, even though $\hat{R}$ values suggest the chains have converged.

5.1.3 Hierarchical Model Fitting of Plasmid Loss Curves

The method of model fitting described above performs reasonably well when fitting to simulated data. However, it assumes that there is a single true value for each parameter in a given strain. In reality one would expect that there is a distribution of behaviours possible for each strain and as such there is a distribution, rather than point value that each parameter can take. Further, one could expect that some of the parameters show similarities between strains and some are unique to the strain. For example, the plasmid loss probability λ is, assuming no partitioning mechanisms, predominantly defined by the plasmid copy number. As such, one would assume

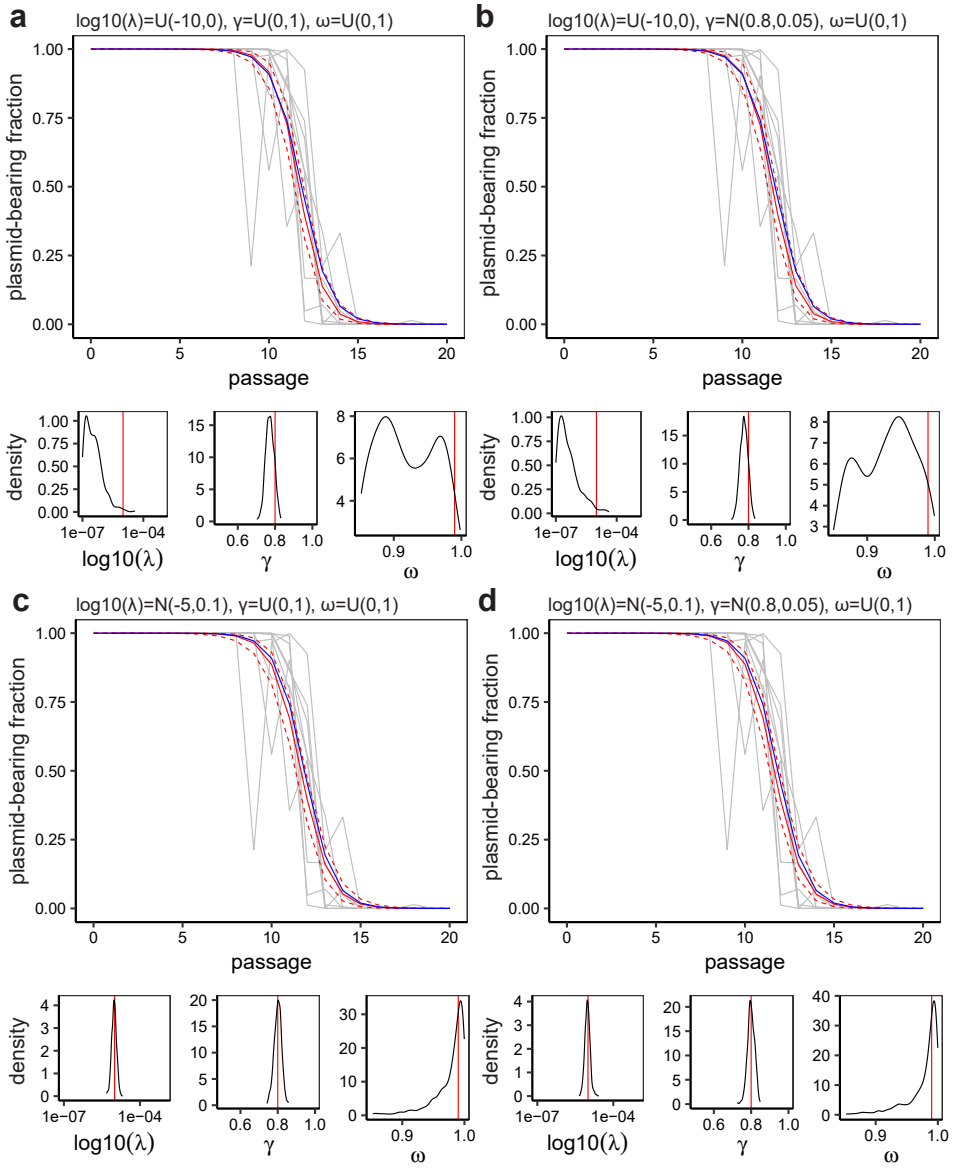

Figure 5.8: Bayesian fitting of simulated data in which there is PSK. The priors used for the fitting are shown above each panel. a) With very broad uniform priors or b) with a tight prior for γ, estimation of λ and ω is very poor. c) With a tight prior on λ, the estimation of ω is improved. d) The addition of a tight prior on γ, as before, yields little benefit. Grey lines show each of the 9 simulated trajectories, the blue line shows the true curve, the solid red line shows the median model fit and the dashed red lines show the 95% confidence intervals. Simulated data produced with $\lambda = 10^{-5}, \gamma = 0.8, \omega = 0.99, \rho = 0.5, N = 1000$. Other priors for fitting were $\rho = U(0,1)$, though the posteriors are not shown.

that the small fluctuations that occur in copy number will create a distribution of plasmid loss probabilities. Further, plasmids based on the same backbone are likely to have similar copy numbers and, therefore their distributions can be thought of a

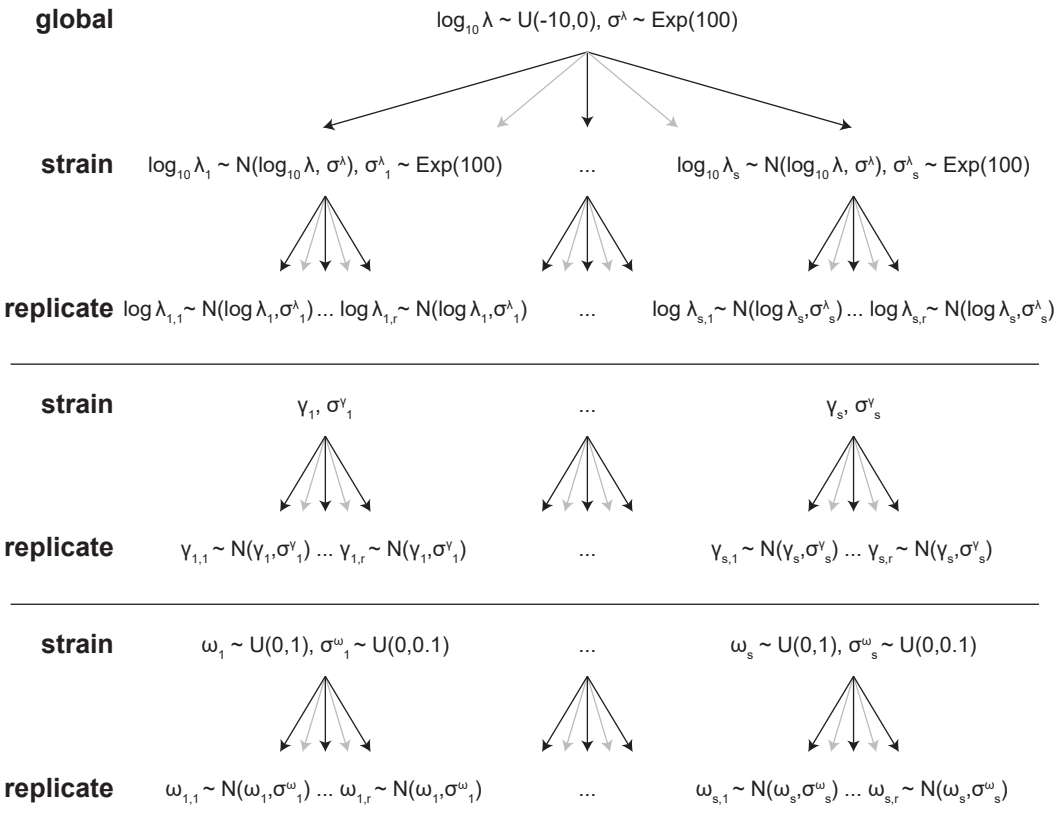

Figure 5.9: Hierarchical model for fitting parameters of the plasmid loss model.

being drawn from a higher level distribution. This is a standard method in Bayesian statistics to model uncertainty in data.

Here I use this concept to build a hierarchical model for the three plasmid loss model parameters, Figure 5.9. As suggested above, the log of the plasmid loss probability λ has a global distribution for all strains with plasmids based on the same backbone. The posterior distribution for this global λ is determined by sampling from a uniform prior, $U(-10,0)$ and the posterior distribution for the standard deviation, σ^{λ}, in the global λ is sampled from an exponential prior, $Exp(100)$, which produces a shrinkage pressure on the distribution, making the posterior distributions narrower. Below the global level, there is a distribution on λ for each strain, sampled from a normal distribution with a mean sampled from the global distribution on λ and a standard deviation sampled from the global distribution on σ^{λ}. There is a strain level standard deviation on the strain level λ which is sampled from an exponential prior in the same manner as the global one. Finally, at the lowest level, each

replicate of a strain has a posterior distribution on λ which is sampled from a normal distribution with mean sampled from the strain level λ and standard deviation sampled from the strain level standard deviation.

The growth ratio parameter, γ could be modelled in the same way because one could assume that the major effector of burden due to the plasmid is in the strong production of the recombinant protein rather than the plasmid stability mechanisms. This assumption is borne out in the growth rate measurements shown later. However, since the growth rates can be directly measured, the strain level distributions can be directly determined. As such, the replicate level posterior distributions for γ are sampled from a normal distribution with mean sampled from the strain level γ and standard deviation given by the standard deviation of the strain level γ.

The parameter encapsulating the efficacy of the killing mechanisms, ω, cannot be said to be globally linked. This is particularly true when considering the differences between the bacteriocin and TA systems but also applies between different TA systems as they function in distinct ways. As such, each strain has a posterior distribution on ω and standard deviation sampled from uniform priors. Each replicate then has a posterior distribution on ω sampled from a normal distribution with mean and standard deviation sampled from the the strain level ω and standard deviation.

The fitting is performed for all data for all strains concurrently. As such, the log likelihood can be written as a sum over all timepoints, t, for all replicate, r, for all strains, s:

$$L(\theta|\mathbf{X}^{+}) = \sum^{s}\sum^{r}\sum^{t}\log\left(P\left(X^{+}_{s,r,t}|\theta\right)\right) \tag{5.11}$$

where θ is the full set of parameters, $\{\lambda_{s,r}, \lambda_s, \sigma_s^{\lambda}, \lambda, \sigma^{\lambda}, \gamma_{s,r}, \gamma_s, \sigma_s^{\gamma}, \omega_{s,r}, \omega_s, \sigma_s^{\omega}, \rho_{s,r}\}$ and:

$$\begin{aligned}
P(X^{+}_{s,r,t}|\theta) &= \text{Beta-Binomial}\left(n_{s,r,t}, \alpha_{s,r,t}, \beta_{s,r,t}\right) \\
\alpha_{s,r,t} &= \phi_{s,r,t}\left(\lambda_{s,r}, \gamma_{s,r}, \omega_{s,r}\right)\left(\frac{1}{\rho_{s,r}} - 1\right) \\
\beta_{s,r,t} &= \frac{\alpha_{s,r,t} - \alpha_{s,r,t}\ \phi_{s,r,t}\left(\lambda_{s,r}, \gamma_{s,r}, \omega_{s,r}\right)}{\phi_{s,r,t}\left(\lambda_{s,r}, \gamma_{s,r}, \omega_{s,r}\right)}
\end{aligned}$$

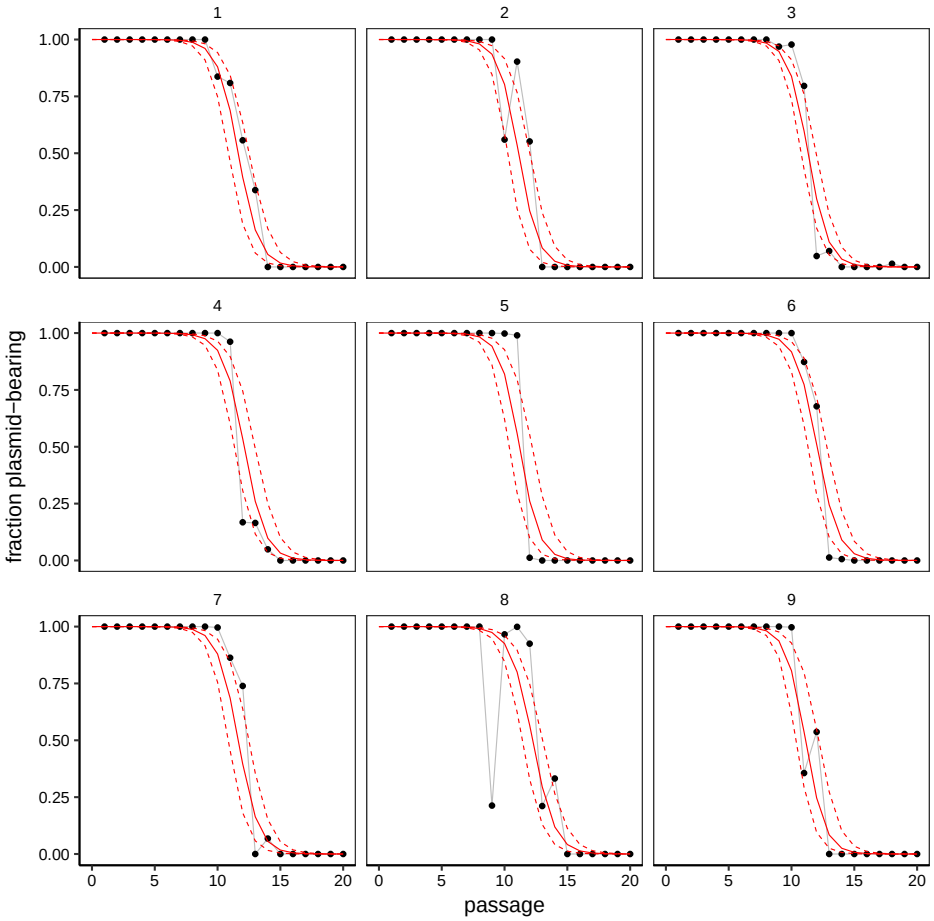

Figure 5.10: Fits of the hierarchical model at the replicate level to simulated plasmid loss curves with PSK. Simulated data are created with parameters $\lambda = 10^{-5}, \gamma = 0.8, \omega = 0.99, \rho = 0.5, N = 1000$. These timecourses are fitted concurrently with simulated data without PSK, $\omega = 0$.

This hierarchical method of fitting can be demonstrated using simulated data of two strains, with and without PSK. The data shown in Figure 5.7 is the data without PSK and the data shown in Figure 5.8 is that with PSK. Both of these data sets are simulated using the same plasmid loss probability, λ, growth rate ratio, γ, and noise, ρ. The difference is in the PSK efficacy parameter, ω. Figure 5.10 shows the ability of this hierarchical method to fit the individual timecourses for the dataset with PSK. Note that each fitted curve is different. This is enabled by the replicate level parameters.

Figure 5.11 shows the model fit to the full set of timecourses for each simulated strain, using the strain level parameter posteriors.

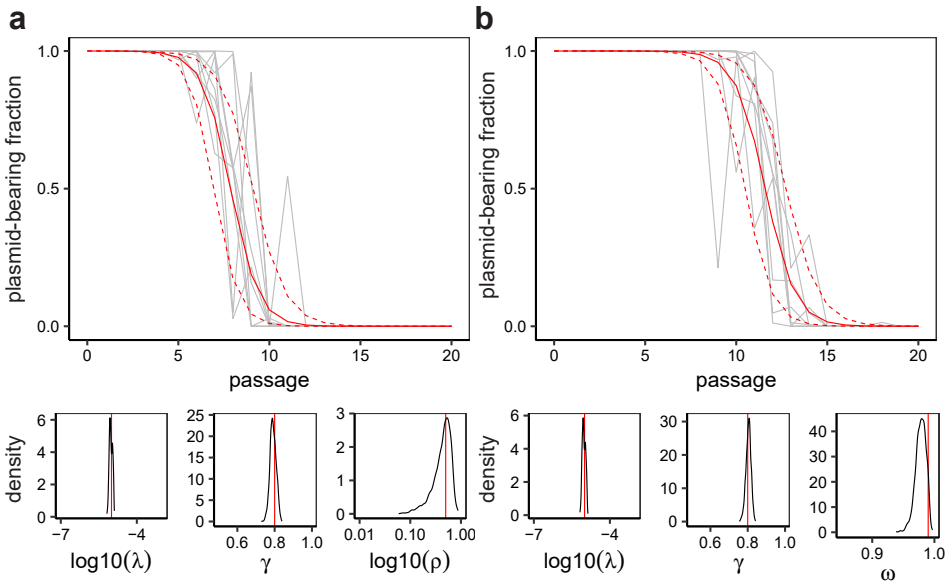

Figure 5.11: Fits of the hierarchical model at the strain level to simulated plasmid loss curves a) without PSK and b) with PSK. Since the ω parameter is not relevant to the fitting of the timecourses without PSK, the posterior distribution of the noise parameter ρ is shown below the plasmid loss curve of (a) instead.

5.1.4 MCMC Model Fitting - Strain Dilution Curves

The models in Equations 5.2 and 5.4 can also be used to fit the data from the strain dilution assays. Whereas, for the plasmid loss curves the data is plasmid-bearing fraction over time, here the initial plasmid-bearing fraction is the varied parameter and plasmid-bearing fraction after 24 hours is the measured variable. Previously the initial population was regarded as entirely plasmid-bearing, $\nu = 0$. Here ν is determined by the initial plasmid-bearing fraction, $\nu = \frac{1}{\phi_0} - 1$. Further, the data is over a single passage which, as discussed previously, is considered to equate to 500 plasmid-free generations, so $\tau = \ln(500)$. As such, the log likelihood is nearly identical to that for the plasmid loss curve fitting but summed over ν rather than t:

$$L(\lambda, \gamma, \omega, \rho | \mathbf{X}^+) = \sum^{r} \sum^{\nu} \log\left(P(x_{r,\nu} | \lambda, \gamma, \omega, \rho)\right) \tag{5.12}$$

The priors for γ are calculated in the same way as in the plasmid loss curve fitting. However, the effects of the plasmid loss parameter, λ, are insignificant in comparison to the effects of the growth difference parameter γ once plasmid-free cells exist, as was discussed previously. As such, λ and consequently the PSK efficacy

parameter ω are removed from the fitting for the TA systems. ω is reintroduced for fitting the bacteriocin system as its effects are not reliant on plasmid loss in this experiment.

5.2 Experimental Determination of Plasmid Stability

A number of bacterial strains were produced in which the stabilising effects of toxin-antitoxin and bacteriocin systems could be experimentally determined, both *in vitro* and *in vivo*. The constructed plasmids, detailed in Section 2.3, include the expression of fluorescent or luminescent proteins that induce a burden on the host strain. Two toxin-antitoxin system, hok-sok and axe-txe, alongside the bacteriocin microcin-V, were tested to determine their stabilising effects. The plasmids were tested in the therapeutically relevant strain *E. coli* Nissle 1917 [178].

5.2.1 Effect of Plasmid Stability Mechanisms on Growth

Initially, plasmids expressing dasher GFP from the strong, constitutive OXB20 promoter in order to produce a burden on plasmid-bearing host cells, were produced. These plasmids were cloned in to a strain of *E. coli* Nissle with the *luxCDABE* cassette and Erythromycin resistance gene in the chromosome (EcN-Lux) [13]. This allows for the growth of the plasmid bearing strain in selective media, to reduce the chance of external contamination while still allowing for plasmid loss.

A microplate reader was used to collect growth data of the plasmid-bearing and plasmid-free strains in LB media with Erythromycin but without Kanamycin. A non-parametric Gaussian process [193] was fitted to the growth curves, Figure 5.12A. This enables the determination of growth rates throughout the growth curve, Figure 5.12B, without having to make the assumptions necessary when fitting standard growth models [194].

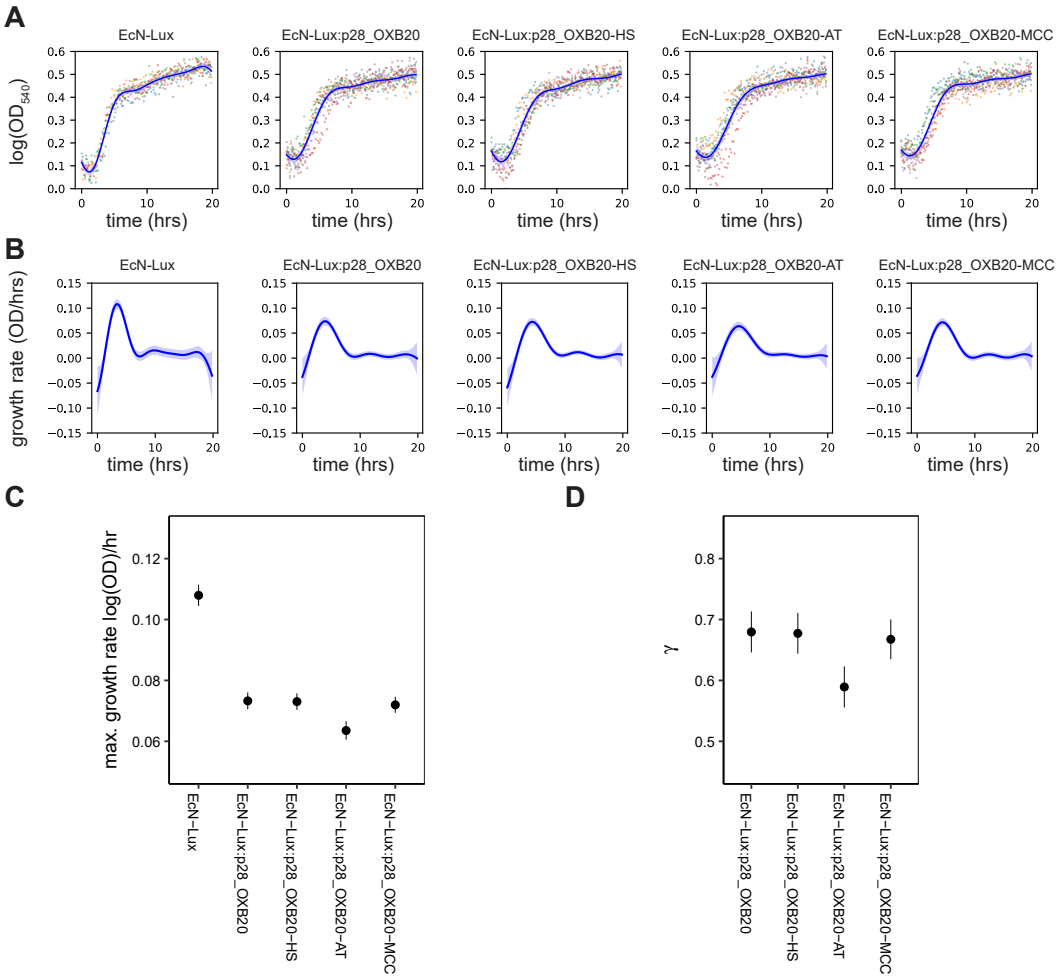

Figure 5.12: Growth curves and model fits for the fluorescent plasmids in EcN-Lux grown in LB media with Erythromycin. (A) Logged optical density measured at 540nm. The points show 6 replicates for each strain with the different colours indicating each replicate. The blue line shows the non-parametric Gaussian process fit with standard error shown by the light blue area around the line. (B) Estimated growth rate as a function of time. (C) Estimated maximal growth rates and (D) growth rate ratios. The points show the mean and the lines show the standard deviation of the mean.

The maximal growth rates of the fluorescent plasmids are shown in Figure 5.12C and the ratio of plasmid-bearing to plasmid-free growth rates in Figure 5.12D. The results show that the burden of expressing GFP from the high copy plasmid results in a growth rate $\sim 70\%$ of the plasmid-free EcN-Lux. Further, the results show that the extra burden of carrying and expressing the plasmid stability mechanisms have minimal impact on growth rates. These results demonstrate the problem for plasmid stability; once the plasmid is dropped, the significantly higher growth in the plasmid-free strain causes the plasmid-bearing population to be outcompeted.

In addition to the fluorescent plasmid, a set of luminescent plasmids constitutively expressing the *LuxCDABE* operon from the phelp promoter [195] were produced for determining the ability of the PSK systems to improve stability *in vivo*. These were transformed in to the wild-type *E. coli* Nissle 1917. The growth rates were determined using the same method as for the fluorescent strains, though the plasmid-bearing strains had to be grown with Kanamycin to ensure retention of the plasmids. This was due to the instability of the control plasmid p24_help-Lux.

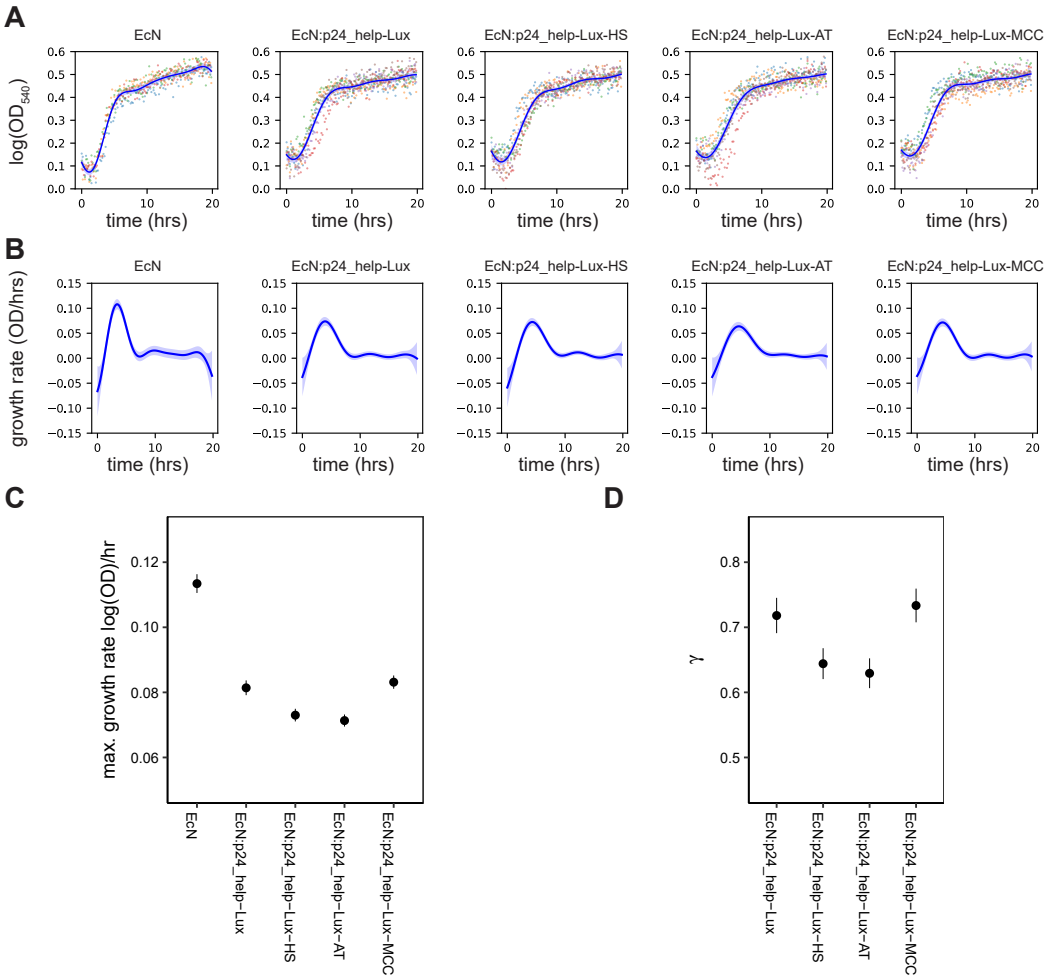

Figure 5.13: Growth curves and model fits for the luminescent plasmids in EcN grown in LB media. (A) Logged optical density measured at 540nm. The points show 6 replicates for each strain with the different colours indicating each replicate. The blue line shows the non-parametric Gaussian process fit with standard error shown by the light blue area around the line. (B) Estimated growth rate as a function of time. (C) Estimated maximal growth rates and (D) growth rate ratios. The points show the mean and the lines show the standard deviation of the mean.

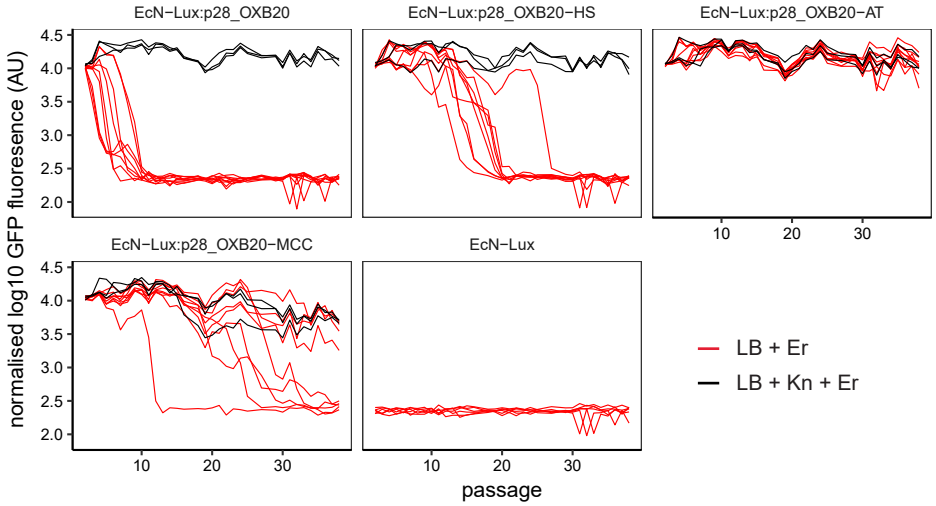

Figure 5.14: GFP fluorescence measurements throughout fluorescent plasmid loss experiment. The black lines (3 replicates) show the trajectories of strains passaged in LB media with Erythromycin and Kanamycin to ensure plasmid maintenance. The red lines (9 replicates for plasmid-bearing strains and 6 replicates for EcN-Lux control) show the trajectories in LB media with Erythromycin but without Kanamycin, allowing for plasmid loss. The data is normalised by negating the mean fluorescence of the EcN-Lux in LB+Er (plasmid-free control) at each passage and scaling so that the lowest recorded measurement is equal to 0.

The growth rate results, Figure 5.13C, show that there is a similar impact on growth rates compared to the fluorescence based plasmids. While the metabolic burden of the *LuxCDABE* operon is thought to be greater than that of GFP, the growth rates are most likely similar due to the lower copy ColE1 origin of replication on the luminescent plasmids compared to the pUC origin on the fluorescent plasmids.

5.2.2 Plasmid Stability Assays *in vitro*

The three stability mechanisms investigated were the *E. coli* native hok/sok type I toxin-antitoxin mechanism, the type II toxin-antitoxin mechanism axe/txe, native to a gram positive bacteria, and the *E. coli* native bacteriocin microcin-V. The fluorescent plasmids containing these stability mechanisms were assayed for their stability in EcN-Lux using a protocol of daily passaging of cultures with samples from each replicate checked for population level fluorescence in a microtitre plate reader, Figure 5.14, and single fluorescence in a flow cytometer, Figures 5.15-5.18.

The flow cytometry data was passed through an automated pipeline to remove

events classified as background debris and cell clumps, returning only single cell events, see Section 2.1.3. The single cell flow cytometry data shows that when plasmid maintenance is enforced through the use of antibiotics the fluorescence levels remain stable across the full 37 passages for all but the microcin-V bearing strain. For the control plasmid and two TA system plasmids, this suggests that the plasmid copy number is stable and the fluorescent output of the plasmid is mutationally stable. It also indicates that samples that are categorised as plasmid-free cells, below, are not the result of loss-of-function mutations leading to a loss of fluorescence. This conclusion is further supported by the population level fluorescence data recorded using a microplate reader remaining stable throughout the experiment, Figure 5.14. However, a reduction in fluorescence is seen in the flow cytometry and microplate reader data for the microcin-V system. The flow cytometry populations are homogeneously fluorescing indicating that, although fluorescence levels are being reduced, it is unlikely that cells being categorised as plasmid-free have undergone a loss-of-function mutation to remove any fluorescence. A reduction in the plasmid copy number is one possible explanation.

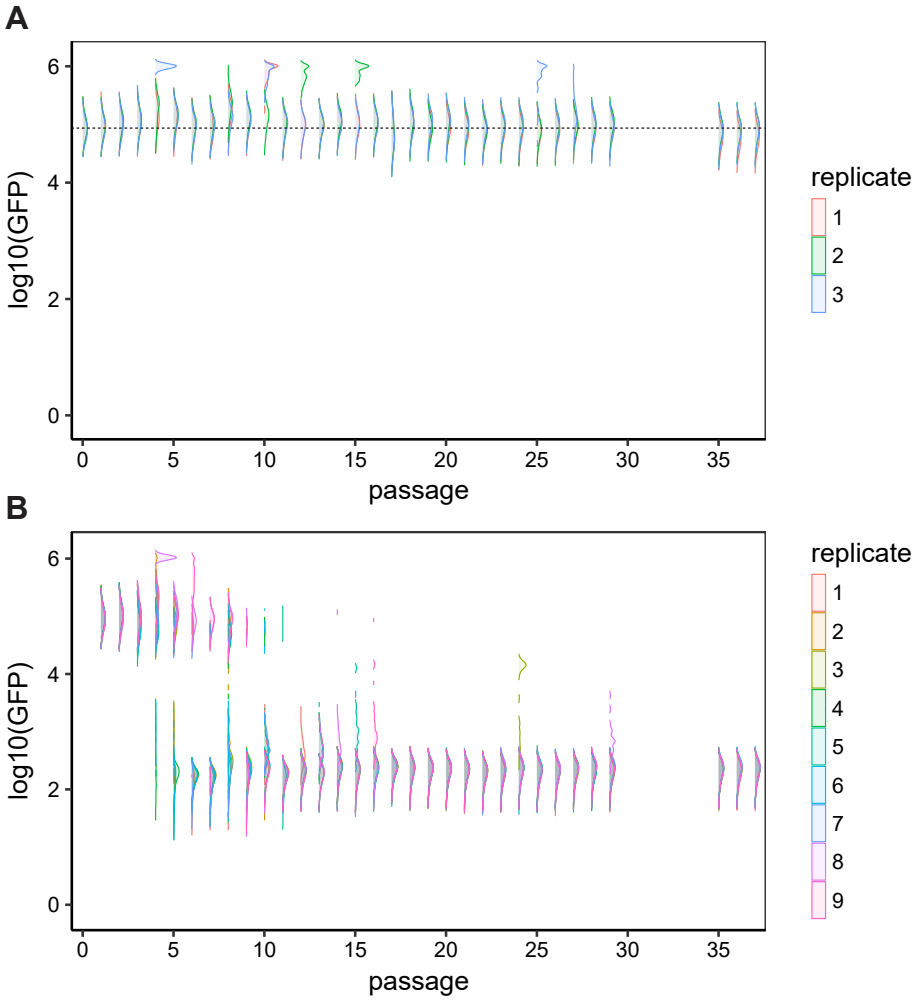

Figure 5.15: Flow cytometry samples of EcN-Lux:p28_OXB20. A) 3 replicates grown in LB media with Erythromycin and Kanamycin, ensuring plasmid maintenance. The black dashed line indicates the median fluorescence at passage 0. B) 9 replicates grown in LB media with Erythromycin only, allowing for plasmid loss.

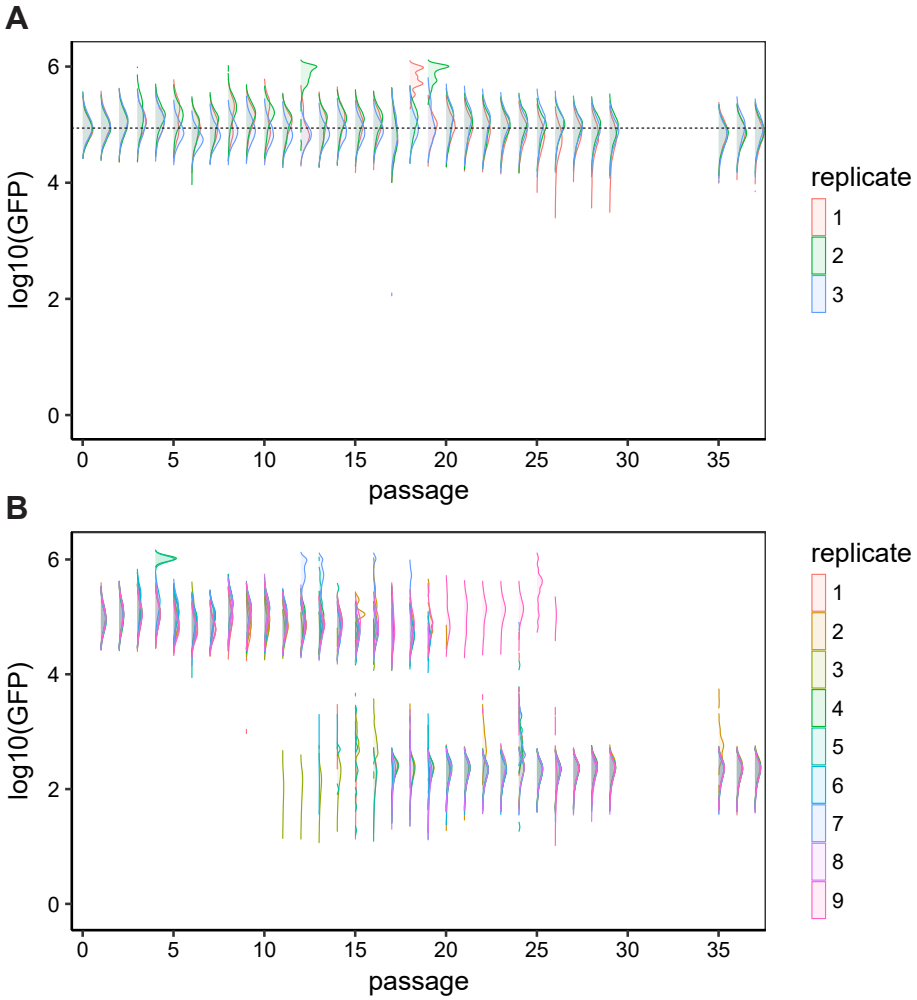

Figure 5.16: Flow cytometry samples of EcN-Lux:p28_OXB20-HS. A) 3 replicates grown in LB media with Erythromycin and Kanamycin, ensuring plasmid maintenance. The black dashed line indicates the median fluorescence at passage 0. B) 9 replicates grown in LB media with Erythromycin only, allowing for plasmid loss.

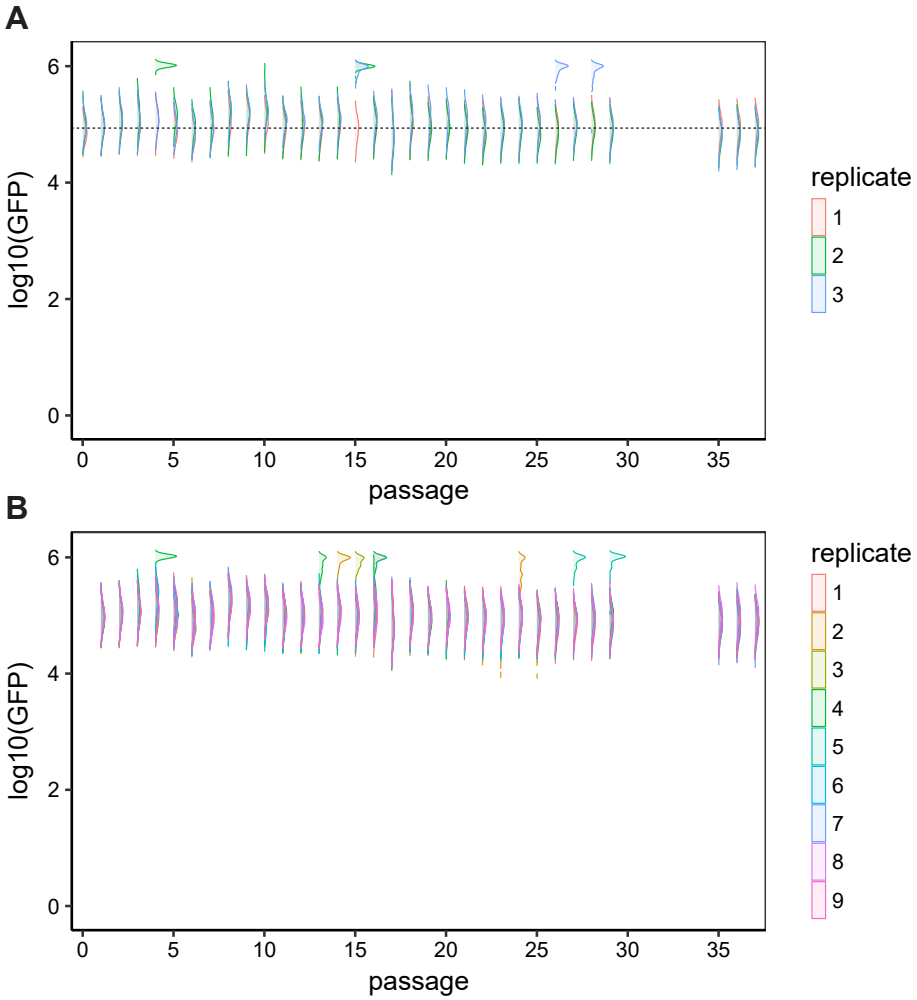

Figure 5.17: Flow cytometry samples of EcN-Lux:p28_OXB20-AT. A) 3 replicates grown in LB media with Erythromycin and Kanamycin, ensuring plasmid maintenance. The black dashed line indicates the median fluorescence at passage 0. B) 9 replicates grown in LB media with Erythromycin only, allowing for plasmid loss.

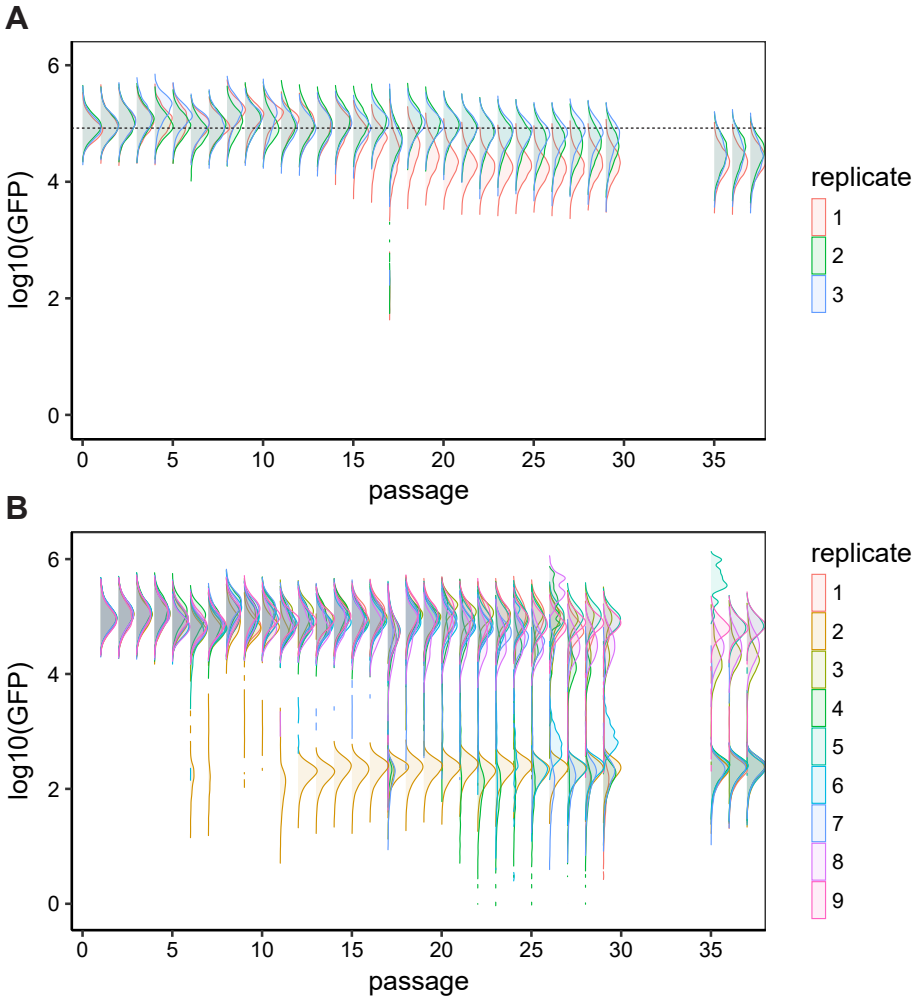

Figure 5.18: Flow cytometry samples of EcN-Lux:p28_OXB20-MCC. A) 3 replicates grown in LB media with Erythromycin and Kanamycin, ensuring plasmid maintenance. The black dashed line indicates the median fluorescence at passage 0. B) 9 replicates grown in LB media with Erythromycin only, allowing for plasmid loss.

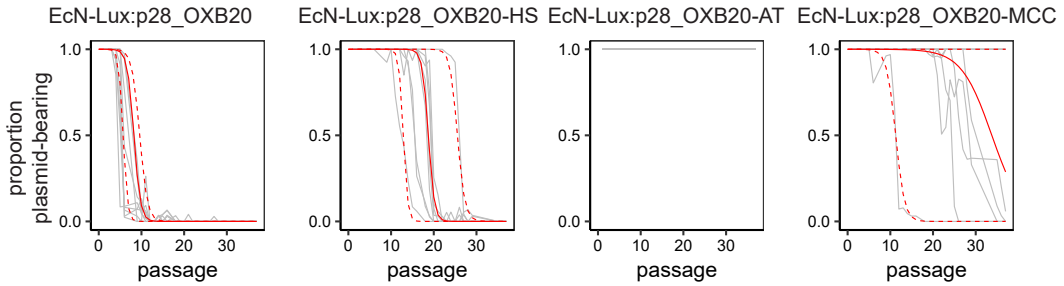

Figure 5.19: Plasmid loss curves for fluorescence based plasmids in EcN-Lux. The grey lines show the trajectories of 9 replicates for each strain. The solid red line shows the model fit and the dashed red lines show the 95% confidence intervals.

The single cell flow cytometry data was then clustered on the GFP fluorescence signal to classify cells as plasmid-free or plasmid-bearing, see Section 2.3.8. This data was used to produce the plasmid loss curves shown in Figure 5.19. For the control plasmid, without a PSK system (EcN-Lux:p28_OXB20), the plasmid-free cells start to become apparent after only $\sim$ 4 passages, with the populations becoming entirely plasmid-free after $\sim$ 12 passages. The results show that the toxin-antitoxin systems improve the stability of the plasmid, though the commonly used hok/sok (EcN-Lux:p28_OXB20-HS) only provides stability for $\sim$ 16 passages, with one replicate beginning to drop as soon as passage 10 and one lasting for nearly 30 passages. Strikingly, axe/txe (EcN-Lux:p28_OXB20-AT) remains stable throughout the experiment, clearly demonstrating its efficacy over the other systems. Because no plasmid loss is observed, the model fitting is essentially non-identifiable. Finally, the results demonstrate that the microcin-V system (EcN-Lux:p28_OXB20-MCC) outperforms the hok/sok system, although the results are heterogeneous, with four of the nine replicates remaining entirely plasmid-bearing for the length of the experiment and in five of the replicates the plasmid-free population takes over.

The posterior of the killing parameter, ω, produced from the model fitting, Figure 5.20, show that even though the hok/sok system has a high killing efficacy it still loses out to competition from the few plasmid-free cells that escape the toxin. On the other hand, even though the bacteriocin system seems to produce a higher plasmid loss rate, perhaps due to a reduction in copy number to compensate for

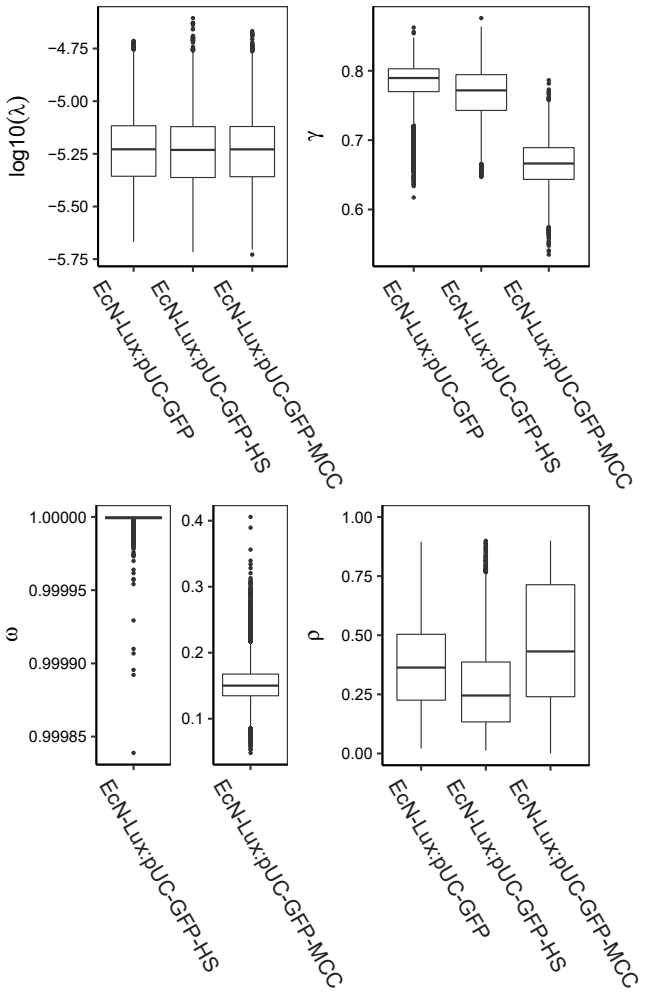

Figure 5.20: Model posteriors for fluorescence based plasmids in EcN-Lux.

increased burden, it needs a far lower killing efficacy to produce better results than hok/sok. The posteriors for the control strain, EcN-Lux:p28_OXB20, show tight control of growth rate but a very wide distribution for the plasmid loss probability, λ. This again may indicate that the pUC based plasmid, with no control on plasmid replication, may be changing copy number to compensate for burden.

Plasmid-stability for the luminescent p24_help-Lux plasmids were then determined using the same growth protocol as for the p28_OXB20-*X* strains. However, measurements could only be carried out using population luminescence in EcN rather than single cell fluorescence in EcN-Lux. This means that we are unable to fit the model to the data as it is not possible to determine the plasmid-bearing and

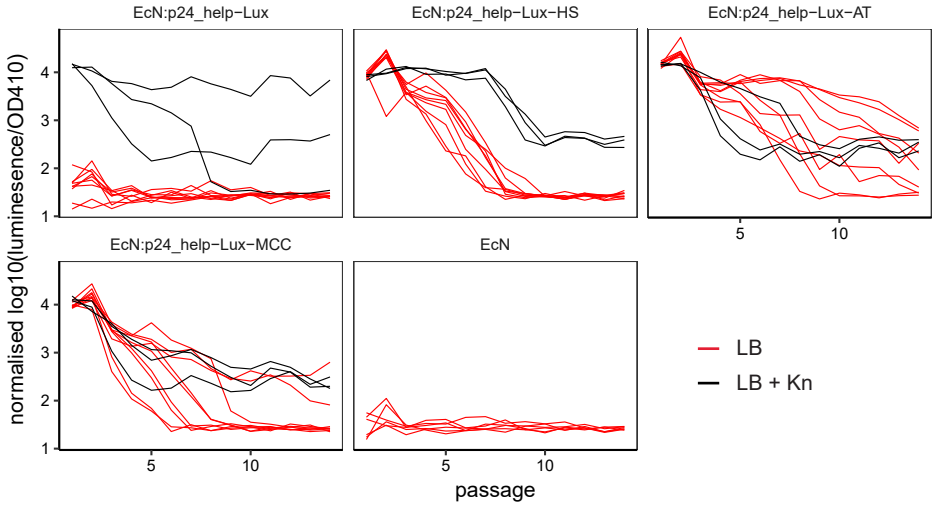

Figure 5.21: Luminescence measurements throughout luminescent plasmid loss experiment. The black lines (3 replicates) show the trajectories of strains passaged in LB media with Kanamycin to ensure plasmid maintenance. The red lines (9 replicates for plasmid-bearing strains and 6 replicates for EcN control) show the trajectories in LB media but without Kanamycin, allowing for plasmid loss. The data is normalised by negating the mean luminescence of the EcN in LB (plasmid-free control) at each passage and scaling so that the lowest recorded measurement is equal to 0.

plasmid-free population proportions. Figure 5.21 shows the population luminescence at each passage measure using a microplate reader. The control population (EcN:p24_help-Lux) had almost stopped luminescing by the end of the first passage. The two TA strains and the microcin-V carrying strain perform better than the control and rank in the same order as for the fluorescent plasmids. However, the luminescence loss occurs faster and even affects the axe/txe bearing strain. Even the samples in which the media contained Kanamycin, enforcing plasmid maintenance, showed reductions in luminescence and in one of the EcN:p24_help-Lux replicates luminescence was completely lost. Unfortunately, as it wasn't possible to record single cell data it cannot be said whether this is a general reduction in luminescence across the whole population, as was seen with the fluorescent microcin-V plasmid, or whether there are populations with no luminescence and populations with stable luminescence.

If one looks at just the luminescence changes in the Kanamycin controls, Fig-

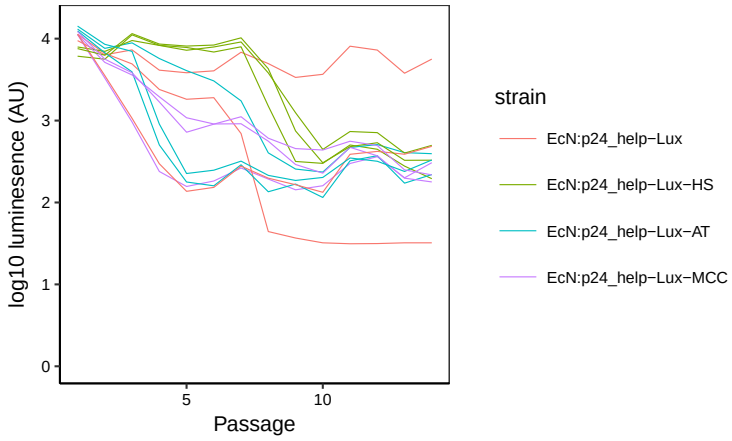

Figure 5.22: Luminescence measurements of Kanamycin control replicates throughout luminescent plasmid loss experiment. Three replicates of each plasmid-bearing strain are shown.

ure 5.22, it is apparent that the populations are tending to a similar lower luminescence state, with the exception of two outliers.

5.2.3 Plasmid Stability Assays in an *in vivo* Tumour Model

Strains of *E. coli* Nissle containing the luminescent plasmids were tested to determine whether the stabilisation mechanisms were functional *in vivo* as well as *in vitro*. The strains were intravenously injected into a mouse tumour model for *in vivo* characterisation [13]. Within 3 days of administration the strains colonised the tumours in the mouse flanks, see Figure 5.23A. Luminescence measurements were taken every day using an *in vivo* imaging system (IVIS) to determine the location and stability of luminescence production from the strains (Figure 5.23).

Seven days after the injection the tumours were excised, homogenised and colony counts were performed in selective and non-selective plates in order to determine the plasmid-bearing proportion of the bacteria within the colonised tumours. Figure 5.24 shows that without any stabilising system $\sim 50\%$ of the population had dropped the p24_help-Lux plasmid after intravenous injection and tumour colonisation. In comparison, the p24_help-Lux-AT was significantly stabilised ($P = 0.0013$) with nearly 100% of the population maintaining the plasmids after colonisation. The strain p24_help-Lux-HS also showed significant stability over p24_help-Lux

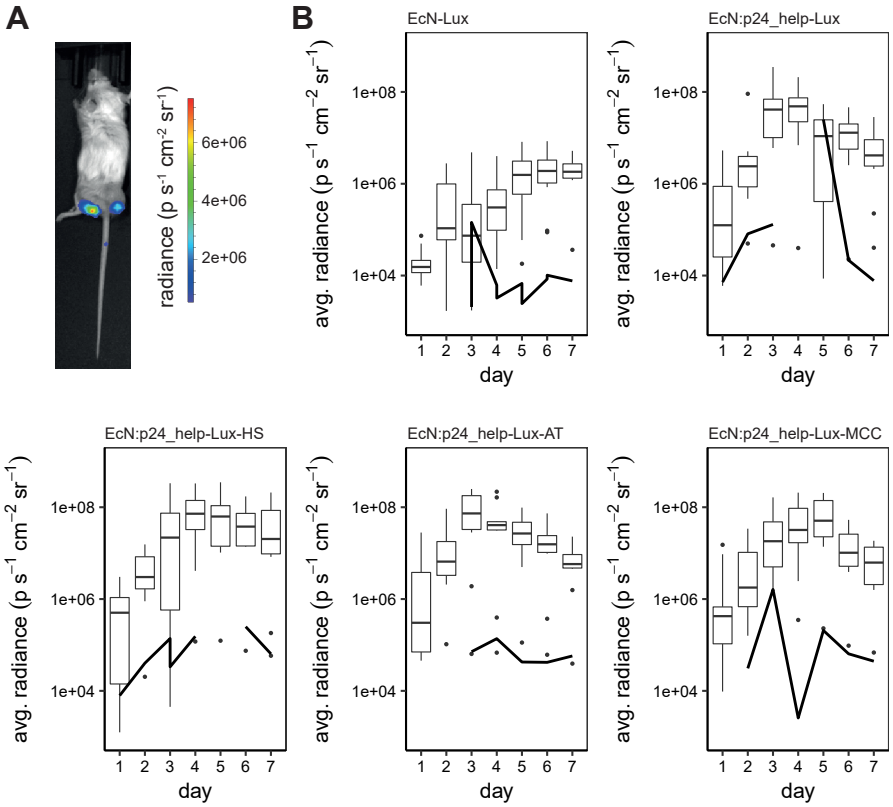

Figure 5.23: Luminescence expression of stabilised plasmids *in vivo*. a) An example of a mouse being imaged in the IVIS system. b) Luminescence measurements of tumours colonised with EcN carrying luminescent plasmids. The control strain EcN-Lux carries a copy of the *luxCDABE* operon on the chromosome. The boxplots show data in measurements of 7 days in which each mouse was implanted with two tumours, one on each flank and for each strain there were 5 mice. Outliers are marked with black points. The black lines show the luminescence measurement of a one blank region on a mouse from each group each day.

control ($P = 0.034$) with $\sim 80\%$ of the population maintaining the plasmids after colonisation. The bacteriocin plasmid p24_help-Lux-MCC did not show a significant improvement over the control though the mean plasmid-bearing population was $\sim 70\%$. There was also no significant difference between the plasmid fraction bearing population in p24_help-Lux-AT and p24_help-Lux-HS.

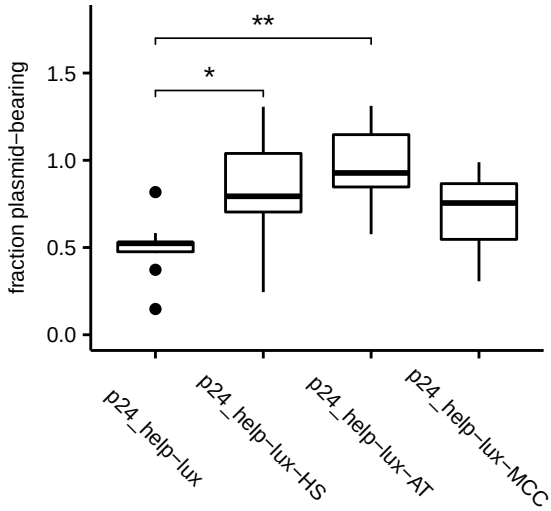

Figure 5.24: Fraction of bacterial population remaining plasmid-bearing 7 days after colonisation, calculated from colony counts on selective and non-selective media performed in triplicate for each tumour. (*P=0.034, **P=0.0013 in the Mann-Whitney U test for p24_help-lux n=9 tumours, p24_help-lux-HS n=7 tumours, p24_help-lux-AT n=8 tumours, p24_help-lux-MCC n=5 tumours)

5.2.4 Restoration of Plasmid-bearing Population with Bacteriocins

As can be seen from the plasmid loss experiments with the fluorescent hok/sok plasmids, once plasmid-free cells arise in the population the toxin-antitoxin system has no way to restore the plasmid-bearing population. Bacteriocins, however, through the secretion of toxin in to the environment provide a general policing of the population. As such, when plasmid-free cells arise or if the culture becomes contaminated the bacteriocin system can push the population back towards entirely plasmid-bearing. To test this, cultures of the fluorescent plasmid based strains were diluted in different ratios with plasmid-free EcN-Lux. The single cell fluorescence was measured using flow cytometry to determine the plasmid-bearing population proportions immediately after the dilutions. The cultures were grown for 24 hours under the same conditions as for the plasmid loss curves, passaged and grown for a further 24 hours. At each passage, single cell fluorescence data was gathered using flow cytometry and the plasmid-bearing population proportions were deter-

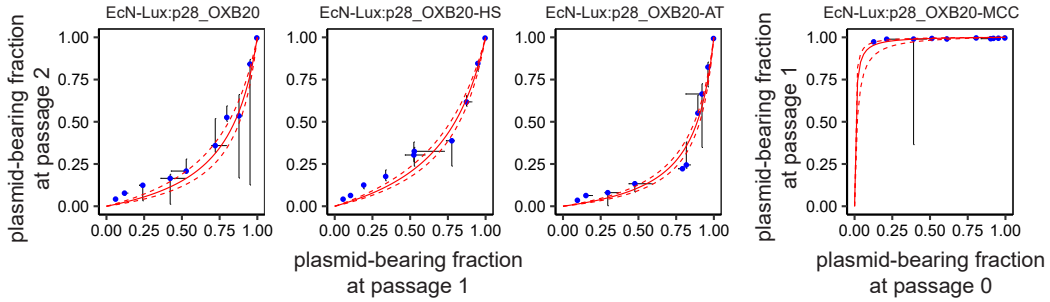

Figure 5.25: Plasmid-bearing strains are diluted with plasmid-free EcN-Lux at different ratios. The dilutions are sampled after 24 hours and show that the faster growth rate of the plasmid-free cells leads to further dilution of the plasmid-bearing population. In the case of EcN-Lux:p28_OXB20-MCC, however, the secreted bacteriocin allows the plasmid-bearing population to out-compete the plasmid-free cells. Blue points show the median of 3 replicates at each initial dilution; the bars show the minimum and maximum plasmid-bearing fraction at each passage. The solid red line shows the model fit and the dashed red lines show the 95% confidence intervals. Each replicate was passaged once initially to allow for the cells to acclimatise to the growth conditions. However, for the EcN-Lux:p28_OXB20-MCC replicates, most plasmid-free cells were dead after 1 passage so the data from passage 0 to passage 1 were used.

mined as previously detailed. Figure 5.25 shows that, when a population of TA plasmid-bearing cells is diluted with plasmid-free EcN-Lux, as stated above the plasmid-free population outgrows the plasmid-bearing population, further diluting it. However, when the EcN-Lux:p28_OXB20-MCC cells are diluted, they quickly kill the plasmid-free cells and restore the population to plasmid-bearing. In fact within 24 hours there are no plasmid-free cells detectable in all but one anomalous sample.

5.3 Discussion

In this Chapter I have outlined the problem of plasmid stability and described one commonly used system and two alternatives for its improvement. ODE models were developed that describe how toxin-antitoxin and bacteriocin systems affect plasmid stability and a Bayesian methodology was established for fitting the models to simulated and experimental data. The PSK systems were tested in the therapeutically relevant *E. coli* Nissle *in vitro* and in an *in vivo* mouse tumour model. Finally, inter-cellular killing by the bacteriocin system was shown to be able a reinstate the

plasmid-bearing population after dilution with plasmid-free bacteria.

The simple Bayesian method developed was able to accurately fit models to systems without PSK systems, even with very limited prior information. This has previously been described as a challenge [192]. The introduction of a PSK system in to the simulated data necessitated the use of prior information for the plasmid loss probability. This demonstrates the correlation between plasmid-loss and post-segregational killing. However, a PSK-free control system can be used to produce this prior information if the assumption that carrying a PSK system doesn't change copy number holds. This assumption is reasonable for plasmids with copy number control but for plasmids without such control it is possible that copy number does change to compensate for increased burden. An unexpected outcome from the Bayesian fitting methodology is the robustness with which growth rate ratios can be determined. This method could conceivably be used to ascertain growth rate differences between strains in competitive cultures rather than using individual growth curve experiments. This may prove incredibly useful when growth rates are dependant on the other strain, such as through cross feeding.

The hierarchical Bayesian fitting method implicitly takes advantage of similarities between strains but also allows for the variability between replicates to be more robustly accounted for. The posterior distribution produced for the plasmid loss probability is more constrained than that produced using the simple approach, given the same level of noise in the simulated data. Although the method, as presented here, used the ability to directly produce strain level growth rate posteriors, this could also be modelled in a way that takes advantage of global similarities.

A set of plasmids were constructed for both fluorescence and luminescence determination of plasmid stability. The fluorescence plasmids enabled the measurement of single cell fluorescence using flow cytometry and, therefore, the determination of plasmid-bearing population proportion without the need for time-consuming and statistically weaker colony-counting methods. The luminescent plasmids allowed the detection of plasmid activity *in vivo* when the plasmid-bearing strains were in the anoxic environment of a tumour, which disqualifies the use of tradi-

tional fluorescent proteins.

Using a microplate reader to measure absorbance and a method for fitting a Gaussian process, growth rates for each of the plasmid bearing strains were determined. This showed that the fluorescent and luminescent reporters expressed on the plasmids produced significant reduction in growth rate. However, the PSK systems did not produce significant extra burden. There was also no significant difference between the fluorescent and luminescent strains. This is unexpected as the expression of luminescence is more costly than that of fluorescence. The promoters driving their expression are rather different, however, with the OXB20 promoter driving GFP expression being particularly strong. Further, the origins of replication on the plasmids are not the same; though they are both high copy, the pUC origin on the fluorescent plasmids is uncontrolled and can produce a far higher copy number than that of the ColE1 origin on the luminescent plasmids.

The fluorescent strains were assayed for plasmid stability using a protocol involving daily passaging and fluorescence measurement using flow cytometry. The data show that the fluorescence burden produces instability resulting in total removal of the plasmid-bearing population after 10 days. All of the PSK mechanisms were able to improve stability with axe/txe greatly outperforming the other systems.

The better performance of axe/txe than of hok/sok is somewhat of a surprise. Firstly, hok/sok is native to *E. coli* and so one might expect it to be more capable in an *E. coli* strain than the non-native axe/txe. Secondly, hok/sok is bactericidal whereas axe/txe is merely bacteristatic. It is reasonable to assume that the plasmid-free persister cells produced from loss of the axe/txe plasmid would at a later stage begin to grow again.

The failure of the microcin-V system to achieve total plasmid stability is also surprising. Changes to the functioning of the plasmid are clear when one considers the change in fluorescence in the selective controls. It is unclear from the data available what led to the reduction in fluorescence. However, the variability in the performance of the microcin-V system suggests that the bacteriocin is not being produced reliably. The regulation of the native bacteriocin promoters is only par-

tially understood but is certainly thought to be influenced by the environment in which the bacteria are growing.

The luminescent strains show a lack of stability of luminescent expression even in the strains grown under selective conditions. This suggests that the loss in fluorescence cannot entirely be explained through plasmid loss. Interestingly, the luminescence of all of the strains grown selectively, with the exception of two outliers, trends towards the same level and plateau, though at different rates. If a sub-population had accrued a loss-of-function mutation that gave it competitive advantage by stopping luminescence production, one would expect the population luminescence levels would trend to the level of the non-luminescent control. As such, it is reasonable to assume that if mutations have arisen, it has been to reduce luminescence rather than stop it. It can also be assumed that the entire population would tend to this luminescence level as strains with higher luminescence would be outcompeted and diluted through passaging. Sequencing of the 5' UTR upstream of the *luxCDABE* cassette could determine if the change in luminescence is due to mutation within the plasmid. Most of the strains grown non-selectively lose luminescence completely, which along with the assumption that complete loss-of-function mutations aren't occurring on the plasmid, suggests that the luminescent plasmids are being lost more rapidly than the fluorescent ones were. The plasmid proves to be so unstable that without a PSK system, luminescence is mostly lost after 24 hours. Indeed the axe/txe system, which was infallible in the fluorescent plasmid, shows total loss of luminescence in 4 of the 9 replicates. The bacteriocin system manages to keep 2 replicates luminescing, beating the hok/sok system again.

Even though the luminescent plasmids show instability in liquid culture, growth conditions *in vivo* are considerably different, with cells growing far more slowly due to poor nutrient conditions within tumours [196]. The *in vitro* imaging results demonstrate that all of the plasmid-bearing strains can colonise tumours and produce luminescence. All of the plasmid-bearing strains perform similarly in terms of luminescence levels, showing slow reduction in luminescence after a peak 3-4 days after administration. The colony-count data, however, shows that the

axe/txe system again performs best in terms of plasmid stability. The bacteriocin shows no significant increase in stability over the non-PSK control. This may be related to the iron-dependant repression of the putative bacteriocin promoters [181]. Environmental iron levels have been implicated in the proliferation of tumours, with iron chelation inhibiting tumour cell proliferation [197]. High iron levels in the tumour microenvironment may lead to the repression of the bacteriocin system. If microcin-V were to be used in a tumour environment in the future, the replacement of the native promoters should be considered.

Finally, it was shown that when challenged with plasmid-free cells at different dilutions, the bacteriocin, as expected, was the only system that enabled the survival of the plasmid-bearing population. This opens up the possibility for the bacteriocin to be used for plasmid stabilisation in environments in which contamination is a risk; though this would require changes to the bacteriocin promoter to make expression more reliable.

5.4 Conclusion

In conclusion, I have produced a new mathematical model of plasmid-loss, integrating post-segregational killing from toxin-antitoxin and bacteriocin systems. I have developed a novel hierarchical Bayesian methodology for fitting this model to experimental data and demonstrated improvement over a standard Bayesian fitting approach. I have experimentally compared three plasmid stabilisation mechanisms in *E. coli* Nissle 1917. This is the first time that axe/txe and mccV have been demonstrated as plasmid stabilisation mechanisms in commensal bacteria. This is also the first demonstration that axe/txe performs better than the current standard mechanism, hok/sok, in any bacterium. I have also shown, for the first time, that axe/txe can stabilise burdensome plasmids in an *in vivo* mouse tumour model with better efficacy than had previously been attained using hok/sok [13].

5.5 Future Work

The toxin-antitoxin and bacteriocin systems demonstrated in this Chapter improved the stability of the plasmids on which they were borne. The model showed that of

the three parameters that affect plasmid stability, only one of them is related to the stability mechanism. However, separating the effects of plasmid loss rate from that of the plasmid stability mechanisms proved difficult. An assumption had to be made that the stability mechanisms had no effect on plasmid copy number, and therefore no effect on plasmid loss rate. This allowed for a control plasmid to be used to infer the loss rates of the stabilised plasmids. A more robust method for the determination of plasmid copy number would be desirable. Quantitative PCR (qPCR) can be used to determine a population average measure of copy number [198, 199]. However, determination of the distribution of plasmid copy number within a population is important [29]. Flow-sorting of cells combine with droplet digital PCR (ddPCR) allows for the determination of copy number within sub-populations that can be chosen based on single cell fluorescence [200]. Along with statistical models, techniques such as these may be able to provide accurate measurements of the distribution of plasmid copy number within a population. Until such methods have been developed and verified, testing the stability mechanisms on plasmids with different origins of replication can provide a semi-quantitative verification of the effect of copy number on plasmid maintenance. As such, the GFP fluorescent plasmid used the plasmid stability assays is being modified with three additional origins of replication: low copy pSC101, medium copy p15A, and the hybrid pRO1600/ColE1 used in the luminescent plasmid. These plasmids will be assayed for plasmid loss using the same method as that used for the plasmids presented in this thesis.

The inconsistency of stabilisation for the bacteriocin system begs many questions. The native promoter was used in this work, but there is only a limited understanding as to its regulation. Experiments using better characterised promoters could potentially improve the consistency of its stabilising effects. In addition, it was noted that the fluorescence levels were reduced in the plasmid-bearing population of the bacteriocin carrying strain. The copy number checks mentioned above could be used to demonstrate whether the fluorescence reduction is due to the extra burden of the bacteriocin causing a reduction in plasmid copy number. Additionally, sequencing checks could be made to determine whether mutation of the regulatory

region is causing the fluorescence reduction.

In addition to the two forms of stability mechanisms explored in this thesis, TA systems and bacteriocins, active partitioning has also been used in previous synthetic biological systems [13]. Naturally, these systems tend to occur on low copy plasmids. Exploration of the efficacy of active partitioning in a range of plasmids and a comparison against the systems presented here would be beneficial to their further use in synthetic biology.

Even if perfect plasmid maintenance is possible, the evolutionary stability of synthetic biological systems encoded on DNA remains a challenge. Particularly for plasmids expressing high amounts of protein, loss-of-function mutants are often heavily selected for. It is possible that TA systems or bacteriocins could provide a solution to a subset of these mutations. Immunity to a toxin can be linked to the continued production of a desired protein by placing the immunity gene in an operon with that protein. Mutations which result in the loss of transcription from the upstream promoter, will also result in loss of immunity, leading to the death of any cells that stop transcribing the operon. Though this is not a full solution to the problem of mutational instability, it is a step that may prove to be useful in the construction of some simple recombinant protein production systems.

Chapter 6

General Conclusions

Synthetic biology has advanced rapidly over the last two decades by combining the tools of molecular biology with the approach and principles of the engineering disciplines. These advances have enabled the further study of biological processes and the development of complex, new-to-biology systems. However, there have only been a few notable examples of synthetic biological systems leaving the laboratory and being deployed in the real world.

In this thesis I have confronted two of the hurdles that currently stand in the way of real world deployment of synthetic biological systems. For both industrial and clinical applications of synthetic biology to advance further, removing the need to use antibiotics and antibiotic resistance genes for circuit stability and contamination control is crucial. The toxin-antitoxin and bacteriocin systems presented here enhance the segregational stability of burdensome plasmids and, therefore, offer an alternative to the use of antibiotic resistance genes. The great challenge for the field now is enhancing the mutational stability of synthetic biological systems. This can perhaps be achieved through the prevention, correction or deletion of mutations. However, a parallel approach that should be explored is the development of new design paradigms that account for random but inevitable mutation events.

The second problem confronted here is that of population control. Division-of-labour is being explored as a method to lessen the burden on individual cells in order to reduce the evolutionary pressure on engineered strains. This requires tools that enable the stable co-existence of multiple strains with differing growth

rates. The SPoCK system, that I present here, for the control of multi-species bacterial populations has the potential to enable predictable, dynamic control over the constituents of a population, while only necessitating the engineering of a single controlling strain. The use of Bayesian modelling, and in particular methods such as approximate Bayesian computation, allows us to explore vast numbers of possible system and experiment designs computationally. As the the systems that we design and build become more complex and control is required at the community level as well the cellular level, using approaches such as these to understand and predict behaviour will become ever more critical.

Summary

In this thesis I have:

- Designed a system to stabilise multi-species bacterial communities using competitive exclusion and dynamic bacteriocin production. This is the first system to use bacteriocins to control community dynamics and the first system to be able to control multi-species communities through the engineering of a single strain.
- Mathematically demonstrated the ability of this system to produce stable or oscillating communities, dependent on easily tunable inputs.
- Demonstrated a theoretical improvement to the robustness of this system through the addition of a self-regulation feedback loop.
- Constructed the first inducible and repressible bacteriocin expression systems.
- Constructed and characterised the first density dependent bacteriocin expression system using a novel Bayesian methodology for parametrising Shea-Ackers promoter dynamics models.
- Undertaken initial experiments with this system in chemostat, demonstrating its ability to extend the presence of the two constituent populations.

- Produced a new mathematical model of plasmid stability, incorporating post-segregational killing from toxin-antitoxin and bacteriocin systems.
- Developed a hierarchical Bayesian methodology for fitting this model and demonstrated its improvement over a standard Bayesian fitting approach.
- Carried out the first characterisation of the axe/txe toxin-antitoxin system and microcin-V bacteriocin in *E. coli* Nissle 1917, *in vitro* and in an *in vivo* mouse tumour model.
- Demonstrated that both of these systems perform better than the current standard system, hok/sok.

Appendix A

Steady State SPoCK Expressions

$$X = \frac{D^2 K_\beta^{-n_\beta} (D(K_X - K_C) + K_C \mu_{X_{max}} - K_X \mu_{C_{max}}) \left(\zeta + K_\beta^{n_\beta}\right)}{k_{\beta_{max}} \omega (D(K_C - K_X) - K_C \mu_{X_{max}})} \tag{A.1}$$

$$\zeta = \left(\frac{\gamma_X k_A \left(\frac{C K_X \mu_{C_{max}}}{D\gamma_C K_C - D\gamma_C K_X - \gamma_C K_C \mu_{X_{max}}} + \frac{D K_X}{D - \mu_{X_{max}}} + S_0 \right)}{D} \right)^{n_\beta} \tag{A.2}$$

$$C = \frac{\gamma_C (D(K_C - K_X) - K_C \mu_{X_{max}}) \left(\xi + \frac{D K_X}{\mu_{X_{max}} - D} - S_0\right)}{K_X \mu_{C_{max}}} \tag{A.3}$$

$$\xi = \frac{D^2 K_\beta^{-n_\beta} (D(K_X - K_C) + K_C \mu_{X_{max}} - K_X \mu_{C_{max}}) \left(\left(\frac{k_A X}{D}\right)^{n_\beta} + K_\beta^{n_\beta}\right)}{\gamma_X k_{\beta_{max}} \omega (D(K_C - K_X) - K_C \mu_{X_{max}})} \tag{A.4}$$

Appendix B

SPoCK Promoter Characterisation Correlograms

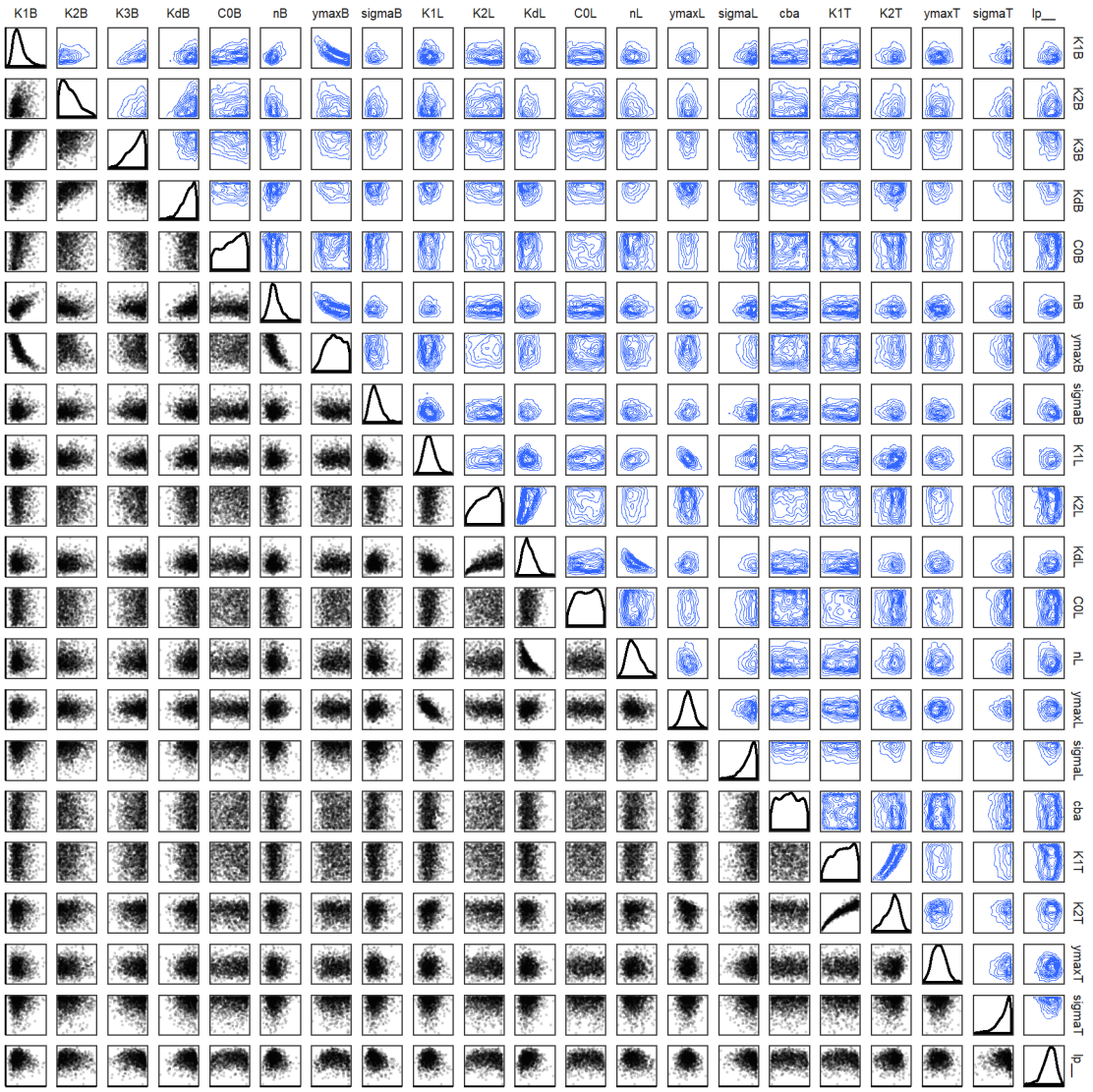

Figure B.1: Correlogram of SPoCK promoter characterisation posteriors. The posterior distribution of each of the parameters in the Shea-Ackers promoter models are shown on the diagonal. The bottom triangle shows the correlation between each pair of parameters for every MCMC sample. The top triangle shows the kernel densities of those correlations. Some parameters are highly correlated, for example K1T and K2T.

Appendix C

SPoCK in Chemostat Results

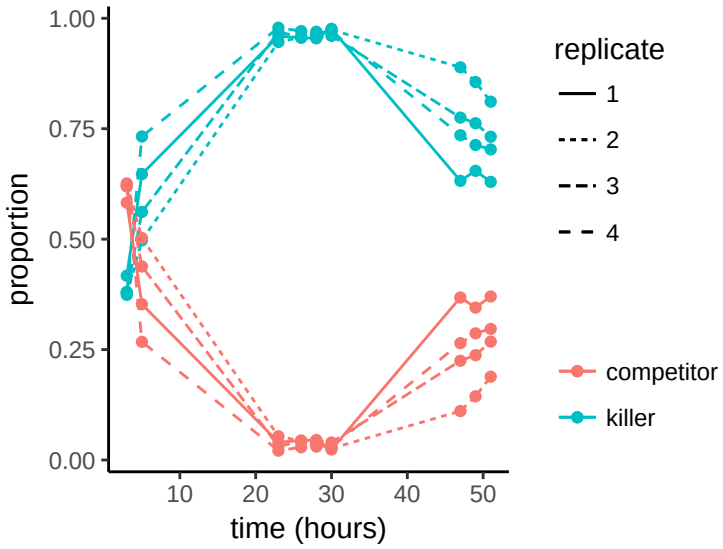

Figure C.1: Bioreactor results of SPoCK in competition. The circles show the points in time at which samples were taken. Replicate 1 and 2 were at $D = 0.1\text{hr}^{-1}$ and replicates 3 and 4 were at $D = 0.05\text{hr}^{-1}$.

Bibliography

[1] Karl Friehs. Plasmid copy number and plasmid stability. In *New trends and developments in biochemical engineering*, pages 47–82. Springer, 2004.

[2] Alec AK Nielsen, Bryan S Der, Jonghyeon Shin, Prashant Vaidyanathan, Vanya Paralanov, Elizabeth A Strychalski, David Ross, Douglas Densmore, and Christopher A Voigt. Genetic circuit design automation. *Science*, 352(6281):aac7341, 2016.

[3] Esteban Martínez-García, Tomás Aparicio, Angel Goñi-Moreno, Sofía Fraile, and Víctor de Lorenzo. Seva 2.0: an update of the standard european vector architecture for de-/re-construction of bacterial functionalities. *Nucleic acids research*, 43(D1):D1183–D1189, 2014.

[4] Jessica G Perez, Jessica C Stark, and Michael C Jewett. Cell-free synthetic biology: engineering beyond the cell. *Cold Spring Harbor perspectives in biology*, 8(12):a023853, 2016.

[5] Baojun Wang, Richard I Kitney, Nicolas Joly, and Martin Buck. Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nature communications*, 2:508, 2011.

[6] Piro Siuti, John Yazbek, and Timothy K Lu. Synthetic circuits integrating logic and memory in living cells. *Nature biotechnology*, 31(5):448, 2013.

[7] Tal Danino, Octavio Mondragón-Palomino, Lev Tsimring, and Jeff Hasty. A synchronized quorum of genetic clocks. *Nature*, 463(7279):326, 2010.

[8] Clementina Dellomonaco, Fabio Fava, and Ramon Gonzalez. The path to next generation biofuels: successes and challenges in the era of synthetic biology. *Microbial Cell Factories*, 9(1):3, 2010.

[9] Chris J Paddon and Jay D Keasling. Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nature Reviews Microbiology*, 12(5):355, 2014.

[10] Marc Valls and Victor De Lorenzo. Exploiting the genetic and biochemical capacities of bacteria for the remediation of heavy metal pollution. *FEMS Microbiology Reviews*, 26(4):327–338, 2002.

[11] Jacob G Bernstein and Edward S Boyden. Optogenetic tools for analyzing the neural circuits of behavior. *Trends in cognitive sciences*, 15(12):592–600, 2011.

[12] Jan Claesen and Michael A Fischbach. Synthetic microbes as drug delivery systems. *ACS synthetic biology*, 4(4):358–364, 2014.

[13] Tal Danino, Arthur Prindle, Gabriel A Kwong, Matthew Skalak, Howard Li, Kaitlin Allen, Jeff Hasty, and Sangeeta N Bhatia. Programmable probiotics for detection of cancer in urine. *Science translational medicine*, 7(289):289ra84–289ra84, 2015.

[14] In Young Hwang, Elvin Koh, Adison Wong, John C March, William E Bentley, Yung Seng Lee, and Matthew Wook Chang. Engineered probiotic escherichia coli can eliminate and prevent pseudomonas aeruginosa gut infection in animal models. *Nature communications*, 8:15028, 2017.

[15] Tanel Ozdemir, Alex J H Fedorec, Tal Danino, and Chris P Barnes. Synthetic biology and engineered live biotherapeutics: Toward increasing system complexity. *Cell Systems*, 7(1):5–16, 2018.

[16] Kenan C Murphy and Kenneth G Campellone. Lambda red-mediated recom-

binogenic engineering of enterohemorrhagic and enteropathogenic e. coli. *BMC molecular biology*, 4(1):11, 2003.

[17] Sheng-Nan Jiang, Seung-Hwan Park, Hee Jung Lee, Jin Hai Zheng, Hyung-Seok Kim, Hee-Seung Bom, Yeongjin Hong, Michael Szardenings, Myung Geun Shin, Sun-Chang Kim, et al. Engineering of bacteria for the visualization of targeted delivery of a cytolytic anticancer agent. *Molecular Therapy*, 21(11):1985–1995, 2013.

[18] Guillaume Achaz, Eric Coissac, Pierre Netter, and Eduardo PC Rocha. Associations between inverted repeats and the structural evolution of bacterial genomes. *Genetics*, 164(4):1279–1289, 2003.

[19] Tomoya Baba, Takeshi Ara, Miki Hasegawa, Yuki Takai, Yoshiko Okumura, Miki Baba, Kirill A Datsenko, Masaru Tomita, Barry L Wanner, and Hirotada Mori. Construction of escherichia coli k-12 in-frame, single-gene knockout mutants: the keio collection. *Molecular systems biology*, 2(1), 2006.

[20] Jack A Bryant, Laura E Sellars, Stephen JW Busby, and David J Lee. Chromosome position effects on gene expression in escherichia coli k-12. *Nucleic acids research*, 42(18):11383–11392, 2014.

[21] Jacob A Englaender, J Andrew Jones, Brady F Cress, Thomas E Kuhlman, Robert J Linhardt, and Mattheos AG Koffas. Effect of genomic integration location on heterologous protein expression and metabolic engineering in e. coli. *ACS synthetic biology*, 6(4):710–720, 2017.

[22] Jeong Wook Lee, Andras Gyorgy, D Ewen Cameron, Nora Pyenson, Kyeong Rok Choi, Jeffrey C Way, Pamela A Silver, Domitilla Del Vecchio, and James J Collins. Creating single-copy genetic circuits. *Molecular cell*, 63(2):329–336, 2016.

[23] Tom Ellis, Tom Adie, and Geoff S Baldwin. Dna assembly for synthetic biology: from parts to pathways and beyond. *Integrative Biology*, 3(2):109–118, 2011.

[24] Arturo Casini, Marko Storch, Geoffrey S Baldwin, and Tom Ellis. Bricks and blueprints: methods and standards for dna assembly. *Nature Reviews Molecular Cell Biology*, 16(9):568, 2015.

[25] Ernesto Andrianantoandro, Subhayu Basu, David K Karig, and Ron Weiss. Synthetic biology: new engineering rules for an emerging discipline. *Molecular systems biology*, 2(1), 2006.

[26] Richard P Novick. Plasmid incompatibility. *Microbiological reviews*, 51(4):381, 1987.

[27] Nileena Velappan, Daniele Sblattero, Leslie Chasteen, Peter Pavlik, and Andrew RM Bradbury. Plasmid incompatibility: more compatible than previously thought? *Protein Engineering, Design and Selection*, 20(7):309–313, 2007.

[28] Oliver Wright, Mihails Delmans, Guy-Bart Stan, and Tom Ellis. Geneguard: a modular plasmid system designed for biosafety. *ACS synthetic biology*, 4(3):307–316, 2014.

[29] David K Summers. The kinetics of plasmid loss. *Trends in biotechnology*, 9(1):273–278, 1991.

[30] Sharon Greenblum, Rogan Carr, and Elhanan Borenstein. Extensive strain-level copy-number variation across human gut microbiome species. *Cell*, 160(4):583–594, 2015.

[31] Jagroop Pandhal and Josselin Noirel. Synthetic microbial ecosystems for biotechnology. *Biotechnology letters*, 36(6):1141–1151, 2014.

[32] Sean C Sleight and Herbert M Sauro. Visualization of evolutionary stability dynamics and competitive fitness of escherichia coli engineered with randomized multigene circuits. *ACS synthetic biology*, 2(9):519–528, 2013.

[33] Jens Kroll, Stefan Klinter, Cornelia Schneider, Isabella Voß, and Alexander Steinbüchel. Plasmid addiction systems: perspectives and applications in biotechnology. *Microbial biotechnology*, 3(6):634–657, 2010.

[34] Bärbel Stecher, Rémy Denzler, Lisa Maier, Florian Bernet, Mandy J Sanders, Derek J Pickard, Manja Barthel, Astrid M Westendorf, Karen A Krogfelt, Alan W Walker, et al. Gut inflammation can boost horizontal gene transfer between pathogenic and commensal enterobacteriaceae. *Proceedings of the National Academy of Sciences*, 109(4):1269–1274, 2012.

[35] Meredith S Wright, Craig Baker-Austin, Angela H Lindell, Ramunas Stepanauskas, Hatch W Stokes, and J Vaun McArthur. Influence of industrial contamination on mobile genetic elements: class 1 integron abundance and gene cassette structure in aquatic bacterial communities. *The ISME journal*, 2(4):417, 2008.

[36] Casey M Theriot, Mark J Koenigsknecht, Paul E Carlson Jr, Gabrielle E Hatton, Adam M Nelson, Bo Li, Gary B Huffnagle, Jun Z Li, and Vincent B Young. Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to clostridium difficile infection. *Nature communications*, 5:3114, 2014.

[37] Jacelyn MS Loh and Thomas Proft. Toxin–antitoxin-stabilized reporter plasmids for biophotonic imaging of group a streptococcus. *Applied microbiology and biotechnology*, 97(22):9737–9745, 2013.

[38] Tao Liu, Jing-yu Chen, Zhong Zheng, Tian-hong Wang, and Guo-Qiang Chen. Construction of highly efficient e. coli expression systems containing low oxygen induced promoter and partition region. *Applied microbiology and biotechnology*, 68(3):346–354, 2005.

[39] Ram Shankar Velur Selvamani, Karl Friehs, and Erwin Flaschel. Extracellular recombinant protein production under continuous culture conditions with

escherichia coli using an alternative plasmid selection mechanism. *Bioprocess and biosystems engineering*, 37(3):401–413, 2014.

[40] Ram Shankar Velur Selvamani, Maurice Telaar, Karl Friehs, and Erwin Flaschel. Antibiotic-free segregational plasmid stabilization in escherichia coli owing to the knockout of triosephosphate isomerase (tpia). *Microbial cell factories*, 13(1):58, 2014.

[41] Marnix H Medema, Axel Trefzer, Andriy Kovalchuk, Marco van den Berg, Ulrike Müller, Wilbert Heijne, Liang Wu, Mohammad T Alam, Catherine M Ronning, William C Nierman, et al. The sequence of a 1.8-mb bacterial linear plasmid reveals a rich evolutionary reservoir of secondary metabolic pathways. *Genome Biology and Evolution*, 2:212–224, 2010.

[42] David K Summers. *The Biology of Plasmids*.

[43] A San Millan, R Peña-Miller, M Toll-Riera, ZV Halbert, AR McLean, BS Cooper, and RC MacLean. Positive selection and compensatory adaptation interact to stabilize non-transmissible plasmids. *Nature communications*, 5:5208, 2014.

[44] Abigail A Salyers, Nadja B Shoemaker, Ann M Stevens, and Lhing-Yew Li. Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. *Microbiological reviews*, 59(4):579–590, 1995.

[45] Eduardo PC Rocha and Antoine Danchin. Base composition bias might result from competition for metabolic resources. *TRENDS in Genetics*, 18(6):291–294, 2002.

[46] Olga Zhaxybayeva. Detection and quantitative assessment of horizontal gene transfer. In *Horizontal Gene Transfer*, pages 195–213. Springer, 2009.

[47] M Victoria Francia, Athanasia Varsaki, M Pilar Garcillán-Barcia, Amparo Latorre, Constantin Drainas, and Fernando de la Cruz. A classification

scheme for mobilization regions of bacterial plasmids. *FEMS microbiology reviews*, 28(1):79–100, 2004.

[48] C Kim, JY Cha, H Yan, SB Vakulenko, and S Mobashery. Hydrolysis of atp by aminoglycoside 3'-phosphotransferases: an unexpected cost to bacteria for harboring an antibiotic resistance enzyme. *The Journal of biological chemistry*, 281(11):6964, 2006.

[49] Marie-Laure Foucault, Patrice Courvalin, and Catherine Grillot-Courvalin. Fitness cost of vana-type vancomycin resistance in methicillin-resistant staphylococcus aureus. *Antimicrobial agents and chemotherapy*, 53(6):2354–2359, 2009.

[50] Florence Depardieu, Isabelle Podglajen, Roland Leclercq, Ekkehard Collatz, and Patrice Courvalin. Modes and modulations of antibiotic resistance gene expression. *Clinical microbiology reviews*, 20(1):79–114, 2007.

[51] Alessandra Carattoli. Resistance plasmid families in enterobacteriaceae. *Antimicrobial agents and chemotherapy*, 53(6):2227–2238, 2009.

[52] LC Cook and GM Dunny. Effects of biofilm growth on plasmid copy number and expression of antibiotic resistance genes in enterococcus faecalis. *Antimicrobial agents and chemotherapy*, 57(4):1850–1856, 2013.

[53] Jasper M Schuurmans, Sacha AFT van Hijum, Jurgen R Piet, Nadine Händel, Jan Smelt, Stanley Brul, and Benno H ter Kuile. Effect of growth rate and selection pressure on rates of transfer of an antibiotic resistance plasmid between e. coli strains. *Plasmid*, 72:1–8, 2014.

[54] Sebastien Gagneux, Clara Davis Long, Peter M Small, Tran Van, Gary K Schoolnik, and Brendan JM Bohannan. The competitive cost of antibiotic resistance in mycobacterium tuberculosis. *Science*, 312(5782):1944–1946, 2006.

[55] Daniel J Mandell, Marc J Lajoie, Michael T Mee, Ryo Takeuchi, Gleb Kuznetsov, Julie E Norville, Christopher J Gregg, Barry L Stoddard, and George M Church. Biocontainment of genetically modified organisms by synthetic protein design. *Nature*, 518(7537):55–60, 2015.

[56] José L Corchero and Antonio Villaverde. Plasmid maintenance in escherichia coli recombinant cultures is dramatically, steadily, and specifically influenced by features of the encoded proteins. *Biotechnology and bioengineering*, 58(6):625–632, 1998.

[57] Juan C Diaz Ricci and Marría Eugenia Hernández. Plasmid effects on escherichia coli metabolism. *Critical reviews in biotechnology*, 20(2):79–108, 2000.

[58] Kenn Gerdes, Martin Howard, and Florian Szardenings. Pushing and pulling in prokaryotic dna segregation. *Cell*, 141(6):927–942, 2010.

[59] Yongfang Li, Alena Dabrazhynetskaya, Brenda Youngren, and Stuart Austin. The role of par proteins in the active segregation of the p1 plasmid. *Molecular microbiology*, 53(1):93–102, 2004.

[60] Simon Ringgaard, Jeroen van Zon, Martin Howard, and Kenn Gerdes. Movement and equipositioning of plasmids by para filament disassembly. *Proceedings of the National Academy of Sciences*, 106(46):19369–19374, 2009.

[61] Robert Ietswaart, Florian Szardenings, Kenn Gerdes, and Martin Howard. Competing para structures space bacterial plasmids equally over the nucleoid. *PLoS computational biology*, 10(12):e1004009, 2014.

[62] P Gayathri, T Fujii, Jakob Møller-Jensen, F van Den Ent, K Namba, and J Löwe. A bipolar spindle of antiparallel parm filaments drives bacterial plasmid segregation. *Science*, 338(6112):1334–1337, 2012.

[63] Jeanne Salje, Pananghat Gayathri, and Jan Löwe. The parmrc system: molec-

ular mechanisms of plasmid segregation by actin-like filaments. *Nature Reviews Microbiology*, 8(10):683, 2010.

[64] David Popp, Weijun Xu, Akihiro Narita, Anthony J Brzoska, Ronald A Skurray, Neville Firth, Umesh Goshdastider, Yuichiro Maéda, Robert C Robinson, and Maria A Schumacher. Structure and filament dynamics of the psk41 actin-like parm protein implications for plasmid dna segregation. *Journal of Biological Chemistry*, 285(13):10130–10140, 2010.

[65] Jakob Møller-Jensen, Jonas Borch, Mette Dam, Rasmus B Jensen, Peter Roepstorff, and Kenn Gerdes. Bacterial mitosis: Parm of plasmid r1 moves plasmid dna by an actin-like insertional polymerization mechanism. *Molecular cell*, 12(6):1477–1487, 2003.

[66] Samuel Million-Weaver and Manel Camps. Mechanisms of plasmid segregation: have multicopy plasmids been overlooked? *Plasmid*, 75:27–36, 2014.

[67] Hanne Ingmer, Christine Miller, and Stanley N Cohen. The repa protein of plasmid psc101 controls escherichia coli cell division through the sos response. *Molecular microbiology*, 42(2):519–526, 2001.

[68] Johan Paulsson and Måns Ehrenberg. Trade-off between segregational stability and metabolic burden: a mathematical model of plasmid cole1 replication control1. *Journal of molecular biology*, 279(1):73–88, 1998.

[69] David K Summers, Christopher WH Beton, and Helen L Withers. Multicopy plasmid instability: the dimer catastrophe hypothesis. *Molecular microbiology*, 8(6):1031–1038, 1993.

[70] Sean D Colloms, Peter Sykora, George Szatmari, and David J Sherratt. Recombination at cole1 cer requires the escherichia coli xerc gene product, a member of the lambda integrase family of site-specific recombinases. *Journal of bacteriology*, 172(12):6973–6980, 1990.

[71] Eleanor L Chant and David K Summers. Indole signalling contributes to the stable maintenance of escherichia coli multicopy plasmids. *Molecular microbiology*, 63(1):35–43, 2007.

[72] Hannah Gaimster and David Summers. Plasmids in the driving seat: The regulatory rna rcd gives plasmid cole1 control over division and growth of its e. coli host. *Plasmid*, 78:59–64, 2015.

[73] CM Field and DK Summers. Multicopy plasmid stability: revisiting the dimer catastrophe. *Journal of theoretical biology*, 291:119–127, 2011.

[74] Shiyin Yao, Donald R Helinski, and Aresa Toukdarian. Localization of the naturally occurring plasmid cole1 at the cell pole. *Journal of bacteriology*, 189(5):1946–1953, 2007.

[75] Joe Pogliano, Thanh Quoc Ho, Zhenping Zhong, and Donald R Helinski. Multicopy plasmids are clustered and localized in escherichia coli. *Proceedings of the National Academy of Sciences*, 98(8):4486–4491, 2001.

[76] Ilaria Cataudella, Ala Trusina, Kim Sneppen, Kenn Gerdes, and Namiko Mitarai. Conditional cooperativity in toxin–antitoxin regulation prevents random toxin activation and promotes fast translational recovery. *Nucleic acids research*, 40(14):6424–6434, 2012.

[77] Elizabeth Diago-Navarro, Ana María Hernández-Arriaga, Slawomir Kubik, Igor Konieczny, and Ramón Díaz-Orejas. Cleavage of the antitoxin of the pard toxin–antitoxin system is determined by the clpap protease and is modulated by the relative ratio of the toxin and the antitoxin. *Plasmid*, 70(1):78–85, 2013.

[78] Nathalie Goeders and Laurence Van Melderen. Toxin-antitoxin systems as multilevel interaction systems. *Toxins*, 6(1):304–324, 2014.

[79] Oliver Wright, Guy-Bart Stan, and Tom Ellis. Building-in biosafety for synthetic biology. *Microbiology*, 159(7):1221–1235, 2013.

[80] Ana J Muñoz-Gómez, Marc Lemonnier, Sandra Santos-Sierra, Alfredo Berzal-Herranz, and Ramón Díaz-Orejas. Rnase/anti-rnase activities of the bacterial pard toxin-antitoxin system. *Journal of bacteriology*, 187(9):3151–3157, 2005.

[81] Hannes Mutschler, Maike Gebhardt, Robert L Shoeman, and Anton Meinhart. A novel mechanism of programmed cell death in bacteria by toxin–antitoxin systems corrupts peptidoglycan synthesis. *PLoS biology*, 9(3):e1001033, 2011.

[82] Simon J Unterholzner, Brigitte Poppenberger, and Wilfried Rozhon. Toxin–antitoxin systems: biology, identification, and application. *Mobile genetic elements*, 3(5):e26219, 2013.

[83] Xiaoxue Wang and Thomas K Wood. Toxin-antitoxin systems influence biofilm and persister cell formation and the general stress response. *Applied and environmental microbiology*, 77(16):5577–5583, 2011.

[84] Paul D Cotter, R Paul Ross, and Colin Hill. Bacteriocinsa viable alternative to antibiotics? *Nature Reviews Microbiology*, 11(2):95, 2013.

[85] Varish Ahmad, Mohd Sajid Khan, Qazi Mohammad Sajid Jamal, Mohammad A Alzohairy, Mohammad A Al Karaawi, and Mughees Uddin Siddiqui. Antimicrobial potential of bacteriocins: in therapy, agriculture and food preservation. *International journal of antimicrobial agents*, 49(1):1–11, 2017.

[86] Shih-Chun Yang, Chih-Hung Lin, Calvin T Sung, and Jia-You Fang. Antibacterial activities of bacteriocins: application in foods and pharmaceuticals. *Frontiers in microbiology*, 5:241, 2014.

[87] Juan Borrero, Yuqing Chen, Gary M Dunny, and Yiannis N Kaznessis. Modified lactic acid bacteria detect and inhibit multiresistant enterococci. *ACS synthetic biology*, 4(3):299–306, 2014.

[88] Saurabh Gupta, Eran E Bram, and Ron Weiss. Genetically programmable pathogen sense and destroy. *ACS synthetic biology*, 2(12):715–23, 2013.

[89] Ana B Campelo, Clara Roces, M Luz Mohedano, Paloma López, Ana Rodríguez, and Beatriz Martínez. A bacteriocin gene cluster able to enhance plasmid maintenance in lactococcus lactis. *Microbial cell factories*, 13(1):77, 2014.

[90] Olivier Borkowski, Carlos Bricio, Michela Murgiano, Brooke Rothschild-Mancinelli, Guy-Bart Stan, and Tom Ellis. Cell-free prediction of protein expression costs for growing cells. *Nature communications*, 9(1):1457, 2018.

[91] Francesca Ceroni, Rhys Algar, Guy-Bart Stan, and Tom Ellis. Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nature methods*, 12(5):415, 2015.

[92] Francisco Dionisio, Ivan Matic, Miroslav Radman, Olivia R Rodrigues, and François Taddei. Plasmids spread very fast in heterogeneous bacterial communities. *Genetics*, 162(4):1525–1532, 2002.

[93] Rembrandt JF Haft, John E Mittler, and Beth Traxler. Competition favours reduced cost of plasmids to host bacteria. *The ISME journal*, 3(7):761, 2009.

[94] Judith E Bouma and Richard E Lenski. Evolution of a bacteria/plasmid association. *Nature*, 335(6188):351, 1988.

[95] Wenying Shou, Sri Ram, and Jose Vilar. Synthetic cooperation in engineered yeast populations. *Proceedings of the National Academy of Sciences*, 104(6):18771882, 2007.

[96] Hyun Jung Kim, James Q. Boedicker, Jang Wook Choi, and Rustem F. Ismagilov. Defined spatial structure stabilizes a synthetic multispecies bacterial community. *Proceedings of the National Academy of Sciences*, 105(47):18188–18193, 8 2008.

[97] Hadeel Majeed, Osnat Gillor, Benjamin Kerr, and Margaret A Riley. Competitive interactions in escherichia coli populations: the role of bacteriocins. *The ISME journal*, 5(1):71, 2011.

[98] Ádám Kerényi, Dóra Bihary, Vittorio Venturi, and Sándor Pongor. Stability of multispecies bacterial communities: signaling networks may stabilize microbiomes. *PLoS One*, 8(3):e57947, 2013.

[99] Hyun-Dong Shin, Shara McClendon, Trinh Vo, and Rachel R Chen. Escherichia coli binary culture engineered for direct fermentation of hemicellulose to a biofuel. *Applied and environmental microbiology*, 76(24):8150–8159, 2010.

[100] Frederick K Balagadd, Hao Song, Jun Ozaki, Cynthia H Collins, Matthew Barnet, Frances H Arnold, Stephen R Quake, and Lingchong You. A synthetic escherichia coli predatorprey ecosystem. *Molecular Systems Biology*, 4(1):187, 8 2008.

[101] Fred Brauer, Carlos Castillo-Chavez, and Carlos Castillo-Chavez. *Mathematical models in population biology and epidemiology*, volume 40. Springer, 2012.

[102] Spencer R Scott, M Omar Din, Philip Bittihn, Liyang Xiong, Lev S Tsimring, and Jeff Hasty. A stabilized microbial ecosystem of self-limiting bacteria using synthetic quorum-regulated lysis. *Nature microbiology*, 2(8):17083, 2017.

[103] Jesus Zaldivar, Jens Nielsen, and Lisbeth Olsson. Fuel ethanol production from lignocellulose: a challenge for metabolic engineering and process integration. *Applied microbiology and biotechnology*, 56(1-2):17–34, 2001.

[104] Jeremy J Minty, Marc E Singer, Scott A Scholz, Chang-Hoon Bae, Jung-Ho Ahn, Clifton E Foster, James C Liao, and Xiaoxia Nina Lin. Design and characterization of synthetic fungal-bacterial consortia for direct production

of isobutanol from cellulosic biomass. *Proceedings of the National Academy of Sciences*, 110(36):14592–14597, 2013.

[105] Madeline A Shea and Gary K Ackers. The or control system of bacteriophage lambda: A physical-chemical model for gene regulation. *Journal of molecular biology*, 181(2):211–230, 1985.

[106] Marc S Sherman and Barak A Cohen. Thermodynamic state ensemble models of cis-regulation. *PLoS computational biology*, 8(3):e1002407, 2012.

[107] Bradley Efron. Bayes' theorem in the 21st century. *Science*, 340(6137):1177–1178, 2013.

[108] Juliane Liepe, Paul Kirk, Sarah Filippi, Tina Toni, Chris P Barnes, and Michael PH Stumpf. A framework for parameter estimation and model selection from experimental data in systems biology using approximate bayesian computation. *Nature protocols*, 9(2):439, 2014.

[109] Walter R Gilks, Sylvia Richardson, and David Spiegelhalter. *Markov chain Monte Carlo in practice*. CRC press, 1995.

[110] Mikael Sunnåker, Alberto Giovanni Busetto, Elina Numminen, Jukka Corander, Matthieu Foll, and Christophe Dessimoz. Approximate bayesian computation. *PLoS computational biology*, 9(1):e1002803, 2013.

[111] Tina Toni, David Welch, Natalja Strelkowa, Andreas Ipsen, and Michael PH Stumpf. Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, 6(31):187–202, 2009.

[112] Miriam Leon, Mae L Woods, Alex J H Fedorec, and Chris P Barnes. A computational method for the investigation of multistable systems and its application to genetic switches. *BMC Systems Biology*, 10(1):130, 2016.

[113] Alex J H Fedorec, Tanel Ozdemir, Anjali Doshi, Luca Rosa, Oscar Velazquez, Tal Danino, and Chris P Barnes. Two new plasmid post-segregational killing mechanisms for the implementation of synthetic gene networks in *E. coli*.

[114] Alex J H Fedorec, Behzad Karkaria, and Chris P Barnes. Dynamic stabilisation of multi-species bacterial consortia using bacteriocins and competitive exclusion.

[115] Christophe Biernacki, Gilles Celeux, and Gérard Govaert. Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE transactions on pattern analysis and machine intelligence*, 22(7):719–725, 2000.

[116] Oksana M Subach, Paula J Cranfill, Michael W Davidson, and Vladislav V Verkhusha. An enhanced monomeric blue fluorescent protein with the high chemical stability of the chromophore. *PloS one*, 6(12):e28674, 2011.

[117] Arthur Prindle, Phillip Samayoa, Ivan Razinkov, Tal Danino, Lev S Tsimring, and Jeff Hasty. A sensing array of radically coupled genetic biopixels. *Nature*, 481(7379):39, 2012.

[118] Kevin D Litcofsky, Raffi B Afeyan, Russell J Krom, Ahmad S Khalil, and James J Collins. Iterative plug-and-play methodology for constructing and modifying synthetic gene networks. *Nature methods*, 9(11):1077, 2012.

[119] L Gilson, HK Mahanty, and R Kolter. Four plasmid genes are required for colicin v synthesis, export, and immunity. *Journal of bacteriology*, 169(6):2466–70, 1987.

[120] Kathryn Geldart, Brittany Forkus, Evelyn McChesney, Madeline McCue, and Yiannis Kaznessis. pmpes: A modular peptide expression system for the delivery of antimicrobial peptides to the site of gastrointestinal infections using probiotics. *Pharmaceuticals*, 9(4):60, 2016.

[121] Kyoung-Hee Choi, Jared B Gaynor, Kimberly G White, Carolina Lopez, Catharine M Bosio, RoxAnn R Karkhoff-Schweizer, and Herbert P Schweizer. A tn7-based broad-range bacterial cloning and expression system. *Nature methods*, 2(6):443, 2005.

[122] David J Warren. Preparation of highly efficient electrocompetent escherichia coli using glycerol/mannitol density step centrifugation. *Analytical biochemistry*, 413(2):206–207, 2011.

[123] Andreas Untergasser, Ioana Cutcutache, Triinu Koressaar, Jian Ye, Brant C Faircloth, Maido Remm, and Steven G Rozen. Primer3new capabilities and interfaces. *Nucleic acids research*, 40(15):e115–e115, 2012.

[124] Kenn Gerdes. The parb (hok/sok) locus of plasmid r1: a general purpose plasmid stabilization system. *Nature Biotechnology*, 6(12):1402, 1988.

[125] Ruth Grady and Finbarr Hayes. Axe–txe, a broad-spectrum proteic toxin–antitoxin system specified by a multidrug-resistant, clinical isolate of enterococcus faecium. *Molecular microbiology*, 47(5):1419–1432, 2003.

[126] Kyoung-Hee Choi and Herbert P Schweizer. mini-tn7 insertion in bacteria with single atttn7 sites: example pseudomonas aeruginosa. *Nature protocols*, 1(1):153, 2006.

[127] Daniel G Gibson, Lei Young, Ray-Yuan Chuang, J Craig Venter, Clyde A Hutchison III, and Hamilton O Smith. Enzymatic assembly of dna molecules up to several hundred kilobases. *Nature methods*, 6(5):343, 2009.

[128] Barry G Hall, Hande Acar, Anna Nandipati, and Miriam Barlow. Growth rates made easy. *Molecular biology and evolution*, 31(1):232–238, 2013.

[129] Aishwariya Sharma. An ultraviolet-sterilization protocol for microtitre plates. *J Exp Microbiol Immunol*, 16:144–147, 2012.

[130] Garrett Hardin. The competitive exclusion principle. *Science*, 131(3409):1292–1297, 1960.

[131] G Evelyn Hutchinson. The paradox of the plankton. *The American Naturalist*, 95(882):137–145, 1961.

[132] Marten Scheffer, Sergio Rinaldi, Jef Huisman, and Franz J Weissing. Why plankton communities have no equilibrium: solutions to the paradox. *Hydrobiologia*, 491(1-3):9–18, 2003.

[133] Mitchell L. Sogin, Hilary G. Morrison, Julie A. Huber, David Mark Welch, Susan M. Huse, Phillip R. Neal, Jesus M. Arrieta, and Gerhard J. Herndl. Microbial diversity in the deep sea and the underexplored rare biosphere. *Proceedings of the National Academy of Sciences*, 103(32):12115–12120, 6 2006.

[134] Michael E. Hibbing, Clay Fuqua, Matthew R. Parsek, and S. Brook Peterson. Bacterial competition: surviving and thriving in the microbial jungle. *Nature Reviews Microbiology*, 8(1):15–25, 1 2009.

[135] Lingchong You, Robert Sidney Cox III, Ron Weiss, and Frances H Arnold. Programmed population control by cell–cell communication and regulated killing. *Nature*, 428(6985):868, 2004.

[136] Calum J. Walsh, Caitriona M. Guinane, Colin Hill, R. Paul Ross, Paul W. OToole, and Paul D. Cotter. In silico identification of bacteriocin gene clusters in the gastrointestinal tract, based on the human microbiome projects reference genome database. *BMC Microbiology*, 15(1):1–11, 9 2015.

[137] Fatima Drissi, Sylvain Buffet, Didier Raoult, and Vicky Merhej. Common occurrence of antibacterial agents in human intestinal microbiota. *Frontiers in Microbiology*, 6:441, 5 2015.

[138] Jinshui Zheng, Michael G. Gnzle, Xiaoxi B. Lin, Lifang Ruan, and Ming Sun. Diversity and dynamics of bacteriocins from human microbiome. *Environmental Microbiology*, 17(6):2133–2143, 6 2015.

[139] Breck A Duerkop, John Varga, Josephine R Chandler, Snow Brook Peterson, Jake P Herman, Mair EA Churchill, Matthew R Parsek, William C Nierman, and E Peter Greenberg. Quorum-sensing control of antibiotic synthesis in burkholderia thailandensis. *Journal of bacteriology*, 191(12):3909–3918, 2009.

[140] Josephine R Chandler, Silja Heilmann, John E Mittler, and E Peter Greenberg. Acyl-homoserine lactone-dependent eavesdropping promotes competition in a laboratory co-culture model. *The ISME journal*, 6(12):2219, 2012.

[141] Karline Soetaert. rootsolve: Nonlinear root finding, equilibrium and steady-state analysis of ordinary differential equations. *R package version*, 1, 2009.

[142] Leonardo Acuña, Gianluca Picariello, Fernando Sesma, Roberto D Morero, and Augusto Bellomio. A new hybrid bacteriocin, ent35–mccv, displays antimicrobial activity against pathogenic gram-positive and gram-negative bacteria. *FEBS Open Bio*, 2(1):12–19, 2012.

[143] Leonardo Acuña, Natalia S Corbalan, Inmaculada C Fernandez-No, Roberto D Morero, Jorge Barros-Velazquez, and Augusto Bellomio. Inhibitory effect of the hybrid bacteriocin ent35-mccv on the growth of escherichia coli and listeria monocytogenes in model and food systems. *Food and Bioprocess Technology*, 8(5):1063–1075, 2015.

[144] Lusine Ghazaryan, Lilit Tonoyan, Ashraf Al Ashhab, M. Ines M. Soares, and Osnat Gillor. The role of stress in colicin regulation. *Archives of Microbiology*, 196(11):753–764, 2014.

[145] Monica I. Abrudan, Fokko Smakman, Ard Jan Grimbergen, Sanne Westhoff, Eric L. Miller, Gilles P. van Wezel, and Daniel E. Rozen. Socially mediated induction and suppression of antibiosis during bacterial coexistence. *Proceedings of the National Academy of Sciences*, 112(35):11054–11059, 9 2015.

[146] Bihter Bayramoglu, David Toubiana, Simon van Vliet, R. Fredrik Inglis, Nadav Shnerb, and Osnat Gillor. Bet-hedging in bacteriocin producing escherichia coli populations: the single cell perspective. *Scientific Reports*, 7:srep42068, 2017.

[147] Rolf Lutz and Hermann Bujard. Independent and tight regulation of transcriptional units in escherichia coli via the lacr/o, the tetr/o and arac/i1-i2 regulatory elements. *Nucleic Acids Research*, 25(6):1203–1210, 1997.

[148] V Solovyev and A Salamov. Automatic annotation of microbial genomes and metagenomic sequences. *Metagenomics and its applications in agriculture, biomedicine and environmental studies*, pages 61–78, 2011.

[149] Katherine Volzing, Konstantinos Biliouris, and Yiannis N Kaznessis. proteon and proteoff, new protein devices that inducibly activate bacterial gene expression. *ACS chemical biology*, 6(10):1107–1116, 2011.

[150] Wai-Leung Ng and Bonnie L Bassler. Bacterial quorum-sensing network architectures. *Annual review of genetics*, 43:197–222, 2009.

[151] Kai Papenfort and Bonnie L Bassler. Quorum sensing signal–response systems in gram-negative bacteria. *Nature Reviews Microbiology*, 14(9):576, 2016.

[152] Paul K Grant, Neil Dalchau, James R Brown, Fernan Federici, Timothy J Rudge, Boyan Yordanov, Om Patange, Andrew Phillips, and Jim Haseloff. Orthogonal intercellular signaling for programmed spatial behavior. *Molecular systems biology*, 12(1):849, 2016.

[153] GS Shadel and TO Baldwin. The vibrio fischeri luxr protein is capable of bidirectional stimulation of transcription and both positive and negative regulation of the luxr gene. *Journal of bacteriology*, 173(2):568–574, 1991.

[154] Luz-Maria Guzman, Dominique Belin, Michael J Carson, and JON Beckwith. Tight regulation, modulation, and high-level expression by vectors con-

taining the arabinose pbad promoter. *Journal of bacteriology*, 177(14):4121–4130, 1995.

[155] Anand Pai, Jaydeep K Srimani, Yu Tanouchi, and Lingchong You. Generic metric to quantify quorum sensing activation dynamics. *ACS synthetic biology*, 3(4):220–227, 2013.

[156] M. Omar Din, Tal Danino, Arthur Prindle, Matt Skalak, Jangir Selimkhanov, Kaitlin Allen, Ellixis Julio, Eta Atolia, Lev S. Tsimring, Sangeeta N. Bhatia, and Jeff Hasty. Synchronized cycles of bacterial lysis for in vivo delivery. *Nature*, 536(7614):81–85, 8 2016.

[157] Jaidip Chatterjee, Carol M Miyamoto, Athina Zouzoulas, B Franz Lang, Nicolas Skouris, and Edward A Meighen. Metr and crp bind to the vibrio harveyi lux promoters and regulate luminescence. *Molecular microbiology*, 46(1):101–111, 2002.

[158] Andrea Y Weiße, Diego A Oyarzún, Vincent Danos, and Peter S Swain. Mechanistic links between cellular trade-offs, gene expression, and growth. *Proceedings of the National Academy of Sciences*, 112(9):E1038–E1047, 2015.

[159] Cordelia PN Rampley, Paul A Davison, Pu Qian, Gail M Preston, C Neil Hunter, Ian P Thompson, Ling Juan Wu, and Wei E Huang. Development of simcells as a novel chassis for functional biosensors. *Scientific reports*, 7(1):7261, 2017.

[160] Tian Tian and Howard M Salis. A predictive biophysical model of translational coupling to coordinate and control protein expression in bacterial operons. *Nucleic acids research*, 43(14):7137–7151, 2015.

[161] Jens Bo Andersen, Claus Sternberg, Lars Kongsbak Poulsen, Sara Petersen Bjørn, Michael Givskov, and Søren Molin. New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Applied and environmental microbiology*, 64(6):2240–2246, 1998.

[162] Gary K Ackers, Alexander D Johnson, and Madeline A Shea. Quantitative model for gene regulation by lambda phage repressor. *Proceedings of the National Academy of Sciences*, 79(4):1129–1133, 1982.

[163] Alvin Tamsir, Jeffrey J Tabor, and Christopher A Voigt. Robust multicellular computing using genetically encoded nor gates and chemical wires. *Nature*, 469(7329):212, 2011.

[164] Bob Carpenter, Andrew Gelman, Matt Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Michael A Brubaker, Jiqiang Guo, Peter Li, Allen Riddell, et al. Stan: A probabilistic programming language. *Journal of Statistical Software*, 20(2):1–37, 2016.

[165] Guy Aidelberg, Benjamin D Towbin, Daphna Rothschild, Erez Dekel, Anat Bren, and Uri Alon. Hierarchy of non-glucose sugars in escherichia coli. *BMC systems biology*, 8(1):133, 2014.

[166] Deborah A Siegele and James C Hu. Gene expression from plasmids containing the arabad promoter at subsaturating inducer concentrations represents mixed populations. *Proceedings of the National Academy of Sciences*, 94(15):8168–8172, 1997.

[167] Chris N Takahashi, Aaron W Miller, Felix Ekness, Maitreya J Dunham, and Eric Klavins. A low cost, customizable turbidostat for use in synthetic circuit characterization. *ACS synthetic biology*, 4(1):32–38, 2014.

[168] K Gerdes, FW Bech, ST Jørgensen, A Løbner-Olesen, PB Rasmussen, T Atlung, L Boe, O Karlstrom, S Molin, and KV Meyenburg. Mechanism of postsegregational killing by the hok gene product of the parb system of plasmid r1 and its homology with the relf gene product of the e. coli relb operon. *The EMBO journal*, 5(8):2023–2029, 1986.

[169] Thomas Franch and Kenn Gerdes. Programmed cell death in bacteria: Translational repression by mrna end-pairing. *Molecular microbiology*, 21(5):1049–1060, 1996.

[170] Thomas Thisted and Kenn Gerdes. Mechanism of post-segregational killing by the hok/sok system of plasmid r1: Sok antisense rna regulates hok gene expression indirectly through the overlapping mok gene. *Journal of molecular biology*, 223(1):41–54, 1992.

[171] Kenn Gerdes, Allan Nielsen, Peter Thorsted, and E Gerhart H Wagner. Mechanism of killer gene activation. antisense rna-dependent rnase iii cleavage ensures rapid turn-over of the stable hok, srnb and pnda effector messenger rnas. *Journal of molecular biology*, 226(3):637–649, 1992.

[172] Kenn Gerdes and E Gerhart H Wagner. Rna antitoxins. *Current opinion in microbiology*, 10(2):117–124, 2007.

[173] Elizabeth M Halvorsen, Julia J Williams, Azra J Bhimani, Emily A Billings, and Paul J Hergenrother. Txe, an endoribonuclease of the enterococcal axe–txe toxin–antitoxin system, cleaves mrna and inhibits protein synthesis. *Microbiology*, 157(2):387–397, 2011.

[174] Sabina Leanti La Rosa, Pat G Casey, Colin Hill, Dzung B Diep, Ingolf F Nes, and Dag A Brede. In vivo assessment of growth and virulence gene expression during commensal and pathogenic lifestyles of luxabcde-tagged enterococcus faecalis strains in murine gastrointestinal and intravenous infection models. *Applied and environmental microbiology*, 79(13):3986–3997, 2013.

[175] Elizabeth M Moritz and Paul J Hergenrother. Toxin–antitoxin systems are ubiquitous and plasmid-encoded in vancomycin-resistant enterococci. *Proceedings of the National Academy of Sciences*, 104(1):311–316, 2007.

[176] Margaret A Riley. Molecular mechanisms of bacteriocin evolution. *Annual review of genetics*, 32(1):255–278, 1998.

[177] R Fredrik Inglis, Bihter Bayramoglu, Osnat Gillor, and Martin Ackermann. The role of bacteriocins as selfish genetic elements. *Biology letters*, 9(3):20121173, 2013.

[178] Lubomir Grozdanov, Carsten Raasch, Jürgen Schulze, Ulrich Sonnenborn, Gerhard Gottschalk, Jörg Hacker, and Ulrich Dobrindt. Analysis of the genome structure of the nonpathogenic probiotic escherichia coli strain nissle 1917. *Journal of bacteriology*, 186(16):5432–5441, 2004.

[179] Martina Sassone-Corsi, Sean-Paul Nuccio, Henry Liu, Dulcemaria Hernandez, Christine T Vu, Amy A Takahashi, Robert A Edwards, and Manuela Raffatellu. Microcins mediate competition among enterobacteriaceae in the inflamed gut. *Nature*, 540(7632):280, 2016.

[180] María F Azpiroz and Magela Laviña. Modular structure of microcin h47 and colicin v. *Antimicrobial agents and chemotherapy*, 51(7):2412–2419, 2007.

[181] Anne E Boyer and Phang C Tai. Characterization of the cvaaand cvi promoters of the colicin v export system: Iron-dependent transcription of cvaa is modulated by downstream sequences. *Journal of bacteriology*, 180(7):1662–1672, 1998.

[182] Li Hong Zhang, Michael J Fath, HK Mahanty, PC Tai, and Roberto Kolter. Genetic analysis of the colicin v secretion pathway. *Genetics*, 141(1):25–32, 1995.

[183] Jaiweon Hwang, Xiaotian Zhong, and Phang C Tai. Interactions of dedicated export membrane proteins of the colicin v secretion system: Cvaa, a member of the membrane fusion protein family, interacts with cvab and tolc. *Journal of bacteriology*, 179(20):6264–6270, 1997.

[184] Jaiweon Hwang, Marina Manuvakhova, and Phang C Tai. Characterization of in-frame proteins encoded by cvaa, an essential gene in the colicin v secretion system: Cvaa* stabilizes cvaa to enhance secretion. *Journal of bacteriology*, 179(3):689–696, 1997.

[185] Karen S Jakes and Alan Finkelstein. The colicin ia receptor, cir, is also the translocator for colicin ia. *Molecular microbiology*, 75(3):567–578, 2010.

[186] Volkmar Braun, Silke I Patzer, and Klaus Hantke. Ton-dependent colicins and microcins: modular design and evolution. *Biochimie*, 84(5-6):365–380, 2002.

[187] Fabien Gérard, Nathalie Pradel, and Long-Fei Wu. Bactericidal activity of colicin v is mediated by an inner membrane protein, sdac, of escherichia coli. *Journal of bacteriology*, 187(6):1945–1950, 2005.

[188] Lars Boe, K Gerdes, and S Molin. Effects of genes exerting growth inhibition and plasmid stability on plasmid maintenance. *Journal of bacteriology*, 169(10):4646–4650, 1987.

[189] L Boe and KV Rasmussen. Suggestions as to quantitative measurements of plasmid loss. *Plasmid*, 36(3):153–159, 1996.

[190] L Boe. Estimation of plasmid loss rates in bacterial populations with a reference to the reproducibility of stability experiments. *Plasmid*, 36(3):161–167, 1996.

[191] Vitaly V Ganusov and Anatoly V Brilkov. Estimating the instability parameters of plasmid-bearing cells. i. chemostat culture. *Journal of theoretical biology*, 219(2):193–205, 2002.

[192] Billy TC Lau, Per Malkus, and Johan Paulsson. New quantitative methods for measuring plasmid loss rates reveal unexpected stability. *Plasmid*, 70(3):353–361, 2013.

[193] Peter S Swain, Keiran Stevenson, Allen Leary, Luis F Montano-Gutierrez, Ivan BN Clark, Jackie Vogel, and Teuta Pilizota. Inferring time derivatives including cell growth rates using gaussian processes. *Nature communications*, 7:13766, 2016.

[194] Boyan Yordanov, Neil Dalchau, Paul K Grant, Michael Pedersen, Stephen Emmott, Jim Haseloff, and Andrew Phillips. A computational method for

automated characterization of genetic components. *ACS synthetic biology*, 3(8):578–588, 2014.

[195] Christian U Riedel, Ian R Monk, Pat G Casey, David Morrissey, Gerald C O'Sullivan, Mark Tangney, Colin Hill, and Cormac GM Gahan. Improved luciferase tagging system for listeria monocytogenes allows real-time monitoring in vivo and in vitro. *Applied and environmental microbiology*, 73(9):3091–3094, 2007.

[196] Tal Danino, Arthur Prindle, Jeff Hasty, and Sangeeta Bhatia. Measuring growth and gene expression dynamics of tumor-targeted s. typhimurium bacteria. *Journal of visualized experiments: JoVE*, (77), 2013.

[197] Suzy V Torti and Frank M Torti. Iron and cancer: more ore to be mined. *Nature reviews Cancer*, 13(5):342, 2013.

[198] Changsoo Lee, Jaai Kim, Seung Gu Shin, and Seokhwan Hwang. Absolute and relative qpcr quantification of plasmid copy number in escherichia coli. *Journal of biotechnology*, 123(3):273–280, 2006.

[199] Mihaela Škulj, Veronika Okršlar, Špela Jalen, Simona Jevševar, Petra Slanc, Borut Štrukelj, and Viktor Menart. Improved determination of plasmid copy number using quantitative real-time pcr for monitoring fermentation processes. *Microbial cell factories*, 7(1):6, 2008.

[200] Michael Jahn, Carsten Vorpahl, Dominique Turkowsky, Martin Lindmeyer, Bruno Buhler, Hauke Harms, and Susann Muller. Accurate determination of plasmid copy number of flow-sorted cells using droplet digital pcr. *Analytical chemistry*, 86(12):5969–5976, 2014.